

DOCUMENT RESUME

ED 237 579

TM 830 850

AUTHOR Rudner, Lawrence M., Ed.
TITLE Testing in Our Schools. Proceedings of the NIE
Invitational Conference on Test Use (Washington,
D.C., June 2, 1980).
INSTITUTION National Inst. of Education (ED), Washington, DC.
PUB DATE 83
NOTE 110p.
PUB TYPE Collected Works - Conference Proceedings (021)
EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Achievement Tests; *Educational Testing; Elementary
Secondary Education; Examiners; *School Districts;
School Surveys; Standardized Tests; *Student
Attitudes; *Teacher Attitudes; *Testing; *Test Use
IDENTIFIERS National Institute of Education; Test Curriculum
Overlap

ABSTRACT

This monograph is designed to produce material for linking testing with instruction. The contents include: (1) Student, Teacher, and Measurement Specialist Attitudes Toward Testing--A Summary of Three Surveys, by Michael D. Beck; (2) Teachers and Testing: A Survey of Knowledge and Attitudes, by James G. Ward; (3) Elementary Teachers and Achievement Tests, by Leslie Salmon-Cox; (4) A Survey of the Use of Various Achievement Tests, by Jennie Yeh, Joan Herman, and Lawrence M. Rudner; (5) Uses of Tests in School Districts, by Mary M. Kennedy; (6) Integrating Assessment with Instruction, by Herbert C. Rudman; and (7) Conference Summary, Conclusions, and Recommendations, by Lawrence M. Rudner. (PN)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

ED237579

TESTING IN OUR SCHOOLS



The National
Institute of
Education
U.S. Department of
Education
Washington, D.C. 20508



TESTING IN OUR SCHOOLS

Proceedings of the NIE Invitational Conference on Test Use

Edited by

Lawrence M. Rudner
Senior Associate
National Institute of Education

U.S. Department of Education
T. H. Bell, Secretary

Office of Educational Research and Improvement
Donald J. Senese, Assistant Secretary

National Institute of Education
Manuel J. Justiz, Director
Shirley Jackson, Associate Director, Teaching and Learning
Jeffry Schiller, Assistant Director, Testing, Assessment and Evaluation

The conference on which this book was based was sponsored by the U.S. Department of Education, National Institute of Education, Testing Assessment and Evaluation Program. The views, opinions, positions, and recommendations within this book are solely those of the authors cited. No official support or endorsement by the National Institute of Education, U.S. Department of Education is intended or should be inferred.

A NOTE TO READERS

Since the 1970's, there has been an increased emphasis and reliance on educational testing, assessment, and evaluation, and a vast number of issues surrounding these subjects has arisen. The issues range from methodological concerns about the validity and reliability of such measurements to political concerns about the appropriate and equitable use of the resulting technical information, and these issues arise at every level of decisionmaking -- from assignment of students in classrooms to Federal policy in appropriating billions of dollars for educational programs.

The Testing, Assessment, and Evaluation (TAE) Division of the National Institute of Education has funded a number of studies on these issues. The Testing Team within TAE supports research on the use of tests in the schools, the impact of testing programs, the relationship of testing to instruction, and new methods for assessing educational outcomes. The Testing Team also sponsors workshops and conferences on testing for educators and the general public, and on the development of materials on testing. The Assessment Team funds the National Assessment of Education Progress (NAEP), sponsors secondary research using the vast NAEP data base and, in addition, reviews State assessment program developments. The Evaluation Team conducts evaluations of urban education programs.

We are pleased to present this monograph on the use of tests and assessment in the classroom. The keen interest in testing and the complexities associated with it make it a priority educational issue, one that must be understood. In response to this need, on June 2, 1980, NIE held a one-day invitational conference on test use. The purpose was to help clarify the issues and to suggest a future agenda. Researchers, most of whose work was supported through NIE contracts and grants, were invited to discuss their past, present, and planned research on test use. They were also invited to serve on a panel responding to questions and recommendations from attendees.

This monograph is the result of the conference, and the Testing Team invites comments concerning its usefulness.

Jeff Schiller
Assistant Director
Testing, Assessment and Evaluation
Program on Teaching and Learning
National Institute of Education
1200 19th Street, N.W.
Washington, D.C. 20208
(202) 254-6271

PREFACE

The presentations at NIE's Conference on Test Use represent the most current research on the use of tests in American schools. Jim Ward, Jennie Yeh, Mary Kennedy, and Herbert Rudman, authors of NIE-funded studies, formed the nucleus of the conference.

Early in 1980, NIE's Testing, Assessment, and Evaluation (TAE) division decided to convene these researchers to present their results to each other and to NIE. As plans for the conference progressed, as coordinator, I invited two more researchers, authors of two recent major studies, Michael Beck and Leslie Salmon-Cox. Because of the widespread interest in test-use research, the test directors from area schools and educational researchers and measurement experts from the East Coast were also invited. In all, over 50 people attended the conference; the participants are listed in the appendix.

Research on test use in the schools was the first line of inquiry developed by NIE's Testing Team. We were interested in obtaining answers to questions such as: how many standardized tests are given annually, and for what purposes? What impact are tests having on students, teachers, and instruction?

Initially, we found that most research on test use was out of date. The last comprehensive works on the use of tests in the schools were published in the mid 1960's (Brim et al., 1964; Brim, 1965; Goslin, 1965). Then, in 1971 Marjorie Kirkland published a comprehensive review of research on the effects of testing. It is interesting to note that her review was reflective of 1960's attitudes toward testing and was fairly positive concerning its effects; her review did not anticipate the widespread criticisms of testing that occurred in the 1970's.

From 1973 to 1977, NIE and other funding agencies supported a major research study on the impact of introducing standardized testing in Ireland (Kellaghan et al., 1980). According to the researchers, "Ireland suggested itself as a locus of the study because that country had a well developed educational system in which, however, very little use was made of standardized testing" (Kellaghan et al., 1980). The findings from the Irish study are very similar to the findings presented in this monograph.

While the Irish study focused primarily on teacher use of test results, Mary Kennedy's study at the Huron Institute explored the use of test information by other decisionmakers, such as administrators and school board members. Her research involved case studies of school districts that have successfully used test and evaluation information.

The aforementioned studies on test use have been complemented by the work of the Center for the Study of Evaluation (CSE), which is funded by NIE under a long-term institutional agreement. The study reported here was first designed and conducted in 1978-79 by CSE. The report for this monograph was the product of a reanalysis of the original CSE test-use data.

The works of Rudman and Ward are the result of two NIE projects designed to help improve the practice of testing in the schools. The projects included such activities as conferences, workshops, and the development of materials on testing for both teachers and parents. In 1978, NIE awarded a grant to the American Federation of Teachers to conduct workshops for teachers and to produce a manual on

how to use standardized tests. The survey results presented by Jim Ward were used in the design of those workshops and conferences.

The final NIE-funded report presented in this monograph is based on a jointly funded study designed to produce materials for linking testing with instruction. NIE supported most of the basic research done by this study, and the Department of Defense Dependent Schools sponsored the development and testing of materials. The contract was awarded to Herbert Rudman, who produced, among several products, a comprehensive review of the literature on test use. Most of the references provided in the back of this monograph were produced by that review.

In the future, NIE will be emphasizing more activities that directly improve test use. Most of these activities will be generated by findings from the test-use studies, and close consultation with State and local test and evaluation directors, test developers, teachers, school administrators, and measurement specialists.

As indicated by this brief overview, NIE's research on test use spans several years and involves a variety of methodological approaches. The studies address test use in diverse ways and focus on uses by different decisionmakers. The results presented at the Test Use Conference reflect the state of the art of research in this area and represent some of the best studies in test use.

TABLE OF CONTENTS

A Note to Readers.....	ii
Preface.....	iii
1. Student, Teacher, and Measurement Specialist Attitudes Toward Testing -- A Summary of Three Surveys by Michael D. Beck, The Psychological Corporation.....	1
2. Teachers and Testing: A Survey of Knowledge and Attitudes by James G. Ward, American Federation of Teachers, AFL-CIO.....	15
3. Elementary Teachers and Achievement Tests by Leslie Salmon-Cox, Learning Research and Development Center.....	25
4. A Survey of the Use of Various Achievement Tests by Jennie Yeh, Center for the Study of Evaluation; Joan Herman, Center of the Study of Evaluation; and Lawrence M. Rudner, National Institute of Education.....	39
5. Uses of Tests in School Districts by Mary M. Kennedy, Huron Institute.....	57
6. Integrating Assessment with Instruction by Herbert C. Rudman, Michigan State University.....	65
7. Conference Summary, Conclusions, and Recommendations by Lawrence M. Rudner, National Institute of Education.....	85
References and Bibliography on Test Use.....	89
List of Participants.....	104

1. STUDENT, TEACHER, AND MEASUREMENT SPECIALIST ATTITUDES TOWARD
STANDARDIZED TESTS -- A SUMMARY OF THREE SURVEYS*

Michael D. Beck, The Psychological Corporation

Most of the data presented in this paper relate to the attitudes of elementary/secondary school teachers and students concerning standardized testing. Additional data provide a comparison of the attitudes of these groups with attitudes predicted or expected of them by a sample of "test experts," specifically, members of the National Council on Measurement in Education. All of these data are based on surveys conducted by Frank Stetz and me in the past 2 years.

The results are a distillation of data drawn from three papers (Beck & Stetz, 1980; Beck & Stetz, 1979; Stetz & Beck, 1979) presented at national meetings. The idea for these studies originated in an earlier pilot study of a similar nature (Stetz & Beck, 1978). The 1979 and 1980 papers are based on two different survey research studies, outlines of which are provided below; additional details concerning the studies are presented in the original papers.

Data Bases

Teacher and Student Attitudes. The sample for this study was composed of all teachers participating in the Spring 1978 standardization of the Metropolitan Achievement Tests. School systems included in the norming group were chosen to yield scores on a sample of students representative of the national school population in terms of geographic region, school system enrollment, socioeconomic status, and public vs. non-public school affiliation. A total of approximately 3500 classroom teachers were included in the norming sample.

The questions used in this study were part of a longer questionnaire designed for the norming program. Teachers were asked to complete the questionnaire as soon as possible after testing was completed. The student portion of the questionnaire was administered orally by the classroom teacher. Teachers recorded the number of hands raised in response to the various questions. Teacher responses were anonymous except for school system identification; completion of the questionnaire was optional.

Teachers were instructed to answer survey questions based on their attitudes toward standardized tests in general, not in terms of the Metropolitan series. It is important to note, however, that teachers knew that test results from this norming program would not be returned to them during the school year. Further, participation in the norming program was not optional at a classroom level. Thus, all teachers had just completed the administration of an experimental version of a test they had not selected or requested to use, and from which they would receive no results. Finally, approximately 75 percent of the teachers had participated in the fall standardization of the tests 6 months earlier, under a similar set of circumstances. Therefore, it is our opinion that this sample did not have a response bias favorable toward standardized tests simply as a result of their participation in the norming program. Further, if any response bias in the sample was likely, it

*Subsequent to the conference, a revised form of this article appeared in Measurement in Education, 12(1), 1981, 1-10.

would have been the tendency to react negatively toward tests. There is, of course, no easy way to evaluate this opinion.

Data in columns 2 and 3 of Table 1 summarize the number of teachers and students responding to the questionnaire. Approximately 95 percent of the teachers participating in the standardization program returned the questionnaire; student responses were received from approximately 88 percent of the norming sample. For purposes of data analysis, the sample was arbitrarily divided according to two variables, grade level, and size and type of school system. Teachers were grouped by grade for Grades K-4, 5-8, and 9-12. For the system size and type variable, the three groups used were teachers in public school systems with fewer than 500 students per grade, public school systems enrolling over 500 students per grade, and all non-public schools. All subsequent analyses were made using these groups. To simplify the presentation of this paper, almost all results are presented in terms of the total sample. Results by size and type of school and by grade level groupings are provided in the original papers (Beck & Stetz, 1979; Stetz & Beck, 1979).

Expectations by "Experts" of Teacher and Student Attitudes. A random sample of 500 National Council on Measurement in Education (NCME) members was selected from the NCME membership list. The teacher and student questionnaires were segmented into three groups; in the questionnaire cover sheet, NCME members were asked to respond as they thought teachers and students would either in Grades K-4, in Grades 5-8, or in Grades 9-12. Assignment of respondents to one of the grade ranges was done systematically.

Materials sent to each NCME member sampled included a cover letter, the questionnaire, and a postpaid return envelope. Approximately 2 weeks later, a followup letter was sent to all sample members, urging their cooperation. To control for response bias due to knowledge of the results of the study, the questionnaire was mailed well in advance of the 1979 annual meetings of NCME and the American Educational Research Association (AERA). The deadline date was also prior to the annual meetings. A total of 284 codable questionnaires were returned, representing a response rate of approximately 57 percent of the original sample. No check was possible on the representativeness of the respondents to the original sample.

Results

Three sets of results will be presented: the attitudes of teachers, the attitudes or reactions of students, and the predictions by "test experts" of the attitudes of teachers and students. This section is organized to present the student

Table 1. Total Number of Teachers, Students and NCME Members Responding by Grade Group

Grades	Teachers	Students	NCME Members
K-4	1516	31,956	103
5-8	1210	28,246	95
9-12	580	11,561	86
Total	3306	71,763	284

reactions, followed by the teacher data; the data from "experts" are included in various portions to provide additional insight into the questions. Not all of the survey results can be presented here due to time and space considerations, so I've attempted to extract what strike me as being the most meaningful, or relevant, or interesting results. The original papers provide additional data.

Student Reactions. Two different sets of questions were used for the student surveys: one for Grades K-4, the other for Grades 5-12. Grades K-4 data will be presented first.

The first questions asked of Grades K-4 students concerned their feelings about standardized achievement tests just before and just after they took such a test. Responses were made using a five-choice smiling-to-frowning-face format. Table 2 summarizes the percent of students indicating each reaction. Students felt somewhat better after the test was completed than before testing. Perhaps this can be interpreted as pleasure or relief that the testing was completed or perhaps to a reduced feeling of anxiety after knowing what a "test" is about; our data do not permit a good analysis of the "why" in this instance. Regardless, it is clear that respondents, on the whole, reacted rather positively toward testing. Collapsing the two smiling and the two frowning faces, from 56 percent (pre) to 72 percent (post) of the sample responded favorably, while 12 percent (pre) and 18 percent (post) reacted negatively.

A further question asked of these students was whether they would like to take another such test next year. Responses were slightly negative: 38 percent said yes, 48 percent no, 14 percent weren't sure or didn't know.

It is interesting to contrast the reactions of these students with the reactions predicted by NCME members. NCME members expected students to react negatively to the test; approximately twice as many of the students had generally positive reactions to the test than was predicted, and about half as many students as predicted responded unfavorably.

Grades 5-12 students were asked a different set of questions concerning their reactions to tests. Results are presented separately for students and NCME members in Table 3. Twice as many students considered teacher-made tests harder than standardized tests. In Grades 9-12, four times as many students considered teacher tests more difficult than standardized tests. Approximately three-quarters of the students considered test questions on standardized tests "generally fair," while

Table 2. Grades K-4 Students' Opinions About Administration of a Standardized Achievement Test¹

	Feelings About the Test				
					
Before the test	37	19	22	9	13
After the test	58	14	10	7	11

¹Numbers represent percents.

Table 3. Comparisons Between Students' Opinions and NCME Members' Impressions of Students' Opinions Toward Standardized Achievement Tests¹
(Grades 5-12)

Questions	Students' Opinions	NCME Members Impressions of Students' Opinions
"Which type of test do you think is usually harder: type of test you just took or the type of test your teachers make up?"		
-Standardized tests	30	62
-Teacher-made tests	59	29
-Don't know	11	9
"The test you just took is supposed to show what you already have learned and what you do not know yet. Do you think the test questions were <u>generally fair</u> ?"		
-Yes	75	53
-No	14	35
-Don't know	11	12
"How do you feel you did on this test?"		
-Well	27	17
-OK/All Right	48	48
-Poorly	7	20
-Don't know	18	16
" <u>In general</u> , how do you feel about the type of test you just took?"		
-Positive	26	38
-Negative	27	31
-Neutral/Unsure	47	31
"Would you be interested in finding out your scores on the test you just took?"		
-Yes	88	76
-No	7	16
-Not Sure	5	8
"Would you like the chance to discuss with your teacher the correct answers to the test you just took?"		
-Yes	43	64
-No	44	24
-Not Sure	13	12
"How many of you get nervous just before you take a test like the one you just took?"	30	65
"How many of you get nervous just before you take the kind of test your teachers make up?"	64	58

¹Numbers represent percents.

about one-seventh of the sample disagreed. Only 7 percent of the sample indicated that they felt they scored poorly on the test they had taken; about 75 percent thought they did "OK, all right," or "well."

General reactions to standardized tests were neutral: one-quarter of the sample reacting positively, one-quarter negatively, and half in between. Most students were interested in receiving their scores on the test, but less than one-half wanted to discuss the test answers with their teacher. (Somewhat more of the Grades 5-8 students than the Grades 9-12 group -- 48 percent versus 31 percent -- expressed interest in discussing the test answers.)

Finally, more than twice as many students expressed nervousness before taking teacher-made tests than before standardized tests. High school students were more nervous prior to teacher-made tests and less anxious prior to standardized tests than were students in Grades 5-8. Comparing these data with a similar question asked of Grades K-4 pupils, almost twice as many of these younger students (56 percent) said they were nervous before taking standardized tests. No comparable question concerning teacher-made tests was asked of this group.

NCME members greatly overestimated the perceived difficulty of standardized tests relative to teacher-made tests. They also overestimated students' negative perceptions of the fairness of the questions on standardized tests. The NCME sample overestimated the percent of students who would like to discuss the test answers and the number of students nervous before standardized tests. In contrast, NCME respondents predicted students' feelings of how they scored on the tests and their general perceptions of such tests rather closely.

In summary, students' reactions at all grades to standardized tests generally ranged from neutral to positive. They consider such tests fair and are interested in how they perform, most believing they score "OK" or better. They express some anxiety about such tests, but less nervousness than they do about teacher-made tests, which they find to be generally more difficult.

Teacher Reactions. Teachers were questioned about their attitudes toward the amount of standardized testing in their school system, their personal use of the results of such instruments, their feelings about the usefulness of tests for various purposes, and their general reactions to such tests.

Table 4 summarizes teachers' responses to the question, "In general, the amount of standardized testing in your school system is...." Overall, 69 percent of the respondents answered "about right." Teachers in the two groups of public school systems responded similarly; respondents in non-public systems were even more satisfied with the amount of testing. Grade-group differences were small, although high school teachers as a group were somewhat more satisfied with the testing programs than were elementary teachers. These data do not support the popularly held notion promoted by test critics that "most" teachers feel too much standardized testing takes place in schools. NCME respondents predicted that a much higher percentage of teachers would consider the amount of testing in their system excessive. Almost 60 percent of the NCME sample thought teachers would reply "too much," while less than 20 percent actually did so.

Teachers were also asked how much they "...personally use standardized achievement test results" in their classrooms. Results are presented in Table 5. Overall, about 10 percent of the respondents make "considerable" use of such results and

Table 4. Comparisons Between Teachers' Opinions and NCME Members' Impressions of Teachers' Opinions of the Amount of Standardized Testing in the Teachers' School System¹

Amount of Testing in your System:	Teachers' Opinions				NCME Members' Impressions of Teachers' Opinions			
	Gr.K-4	5-8	9-12	Total	Gr.K-4	5-8	9-12	Total
Too Great	22	18	16	19	65	55	57	59
About Right	67	72	64	69	35	43	38	39
Too Little	5	6	13	7	0	1	5	2
No Answer	6	4	7	5	0	1	0	0

¹Numbers represent percents.

Table 5. Percent of Teachers Indicating Various Amounts of Personal Use of Standardized Achievement Test Results Compared with Percent of Teachers Predicted by NCME Members

Personal Use of Test Results:	Teachers	NCME Members' Impressions
Considerable	9	7
Some	50	57
Little	31	34
None/Not Applicable	10	1

just under 50 percent make "some" use of test data. High school teachers made significantly less use of results than did primary and elementary teachers. On this question, the NCME sample's perceptions were fairly close to the responses of teachers on this question.

The questionnaire listed eight possible uses of standardized achievement test results. Teachers were asked to indicate whether they personally used tests for these purposes, and to indicate any other uses they made of such data. Across all respondents, an average of four uses were checked or listed. Table 6 summarizes the percent of teachers who indicated they use standardized achievement results for the eight listed purposes. The uses checked most frequently were for "diagnosing strengths and weaknesses" (75 percent), "measuring 'growth'" (66 percent), and "individual student evaluation" (65 percent). The most infrequently checked use was for "reporting to students" (24 percent). Fewer than 200 responses were written in by teachers for this question; most free responses were "for grouping" and "to evaluate funded programs." Teachers in large public systems indicated fewer uses of test data and non-public teachers made substantially more use of test results. Again, high school teachers used tests less than did teachers in other grades. The NCME members' responses yielded a rank order of the various uses made very similar to that of the teachers, although NCME respondents generally predicted significantly fewer uses than were actually indicated by teachers.

Table 6. Percent of Teachers Making Various Uses of Standardized Achievement Test Results in Their Classrooms and Corresponding Percent Predicted by NCME Members

Personally Use Standardized Achievement Test Results For:	Teachers	NCME Members' Impressions
Individual student evaluation	65	63
Diagnosing strengths and weaknesses	74	56
Class evaluation	45	51
Instructional planning	52	36
Evaluation of teaching methods	37	11
Reporting to parents	42	82
Reporting to students	24	50
Measuring "growth"	66	57

The next question surveyed teacher opinions concerning the usefulness of standardized achievement test results for 17 various purposes listed in Table 7. A majority of teachers rated such instruments "useful" for 11 of the 17 purposes listed. For the total sample, the responses can be summarized as follows:

*Over 75 percent of teachers consider such tests useful for:

- measuring educational "growth" of individual students
- detecting system-wide general strengths and weaknesses.

*Between 60 percent and 70 percent of respondents considered such tests useful for:

- reporting to parents
- helping to plan instruction for class groups
- helping to plan instruction for individual students
- measuring the educational status of individual students.

*Fewer than one-third of the teachers considered such tests useful for:

- reporting to newspapers (10 percent)
- helping to evaluate teacher performance (21 percent)
- comparing classes within a school (30 percent).

As the data in Table 7 indicate, NCME members tended to overestimate teachers' dispositions toward test use for "reporting" and "comparison" purposes and to underestimate teachers' use for "instructional planning" functions.

Teachers' general reactions to standardized achievement tests were investigated through 11 semantic differential scales. The scales used and the resulting responses are summarized in Tables 8 and 9. Although the questionnaire used a seven-point response scale, data were collapsed into a three-point scale range for summary purposes. As Table 8 indicates, teachers rated themselves as generally calm, comfortable, and knowledgeable about standardized tests, as interested in them, and supportive of them. The NCME predictions of teacher reactions were far more negative than were actually found. Table 9 results reveal that teachers find standardized achievement tests hard, but generally helpful, useful, and fair; on the average, respondents were fairly neutral on the unbiased-biased and valid-invalid continuums. Teachers generally rated standardized tests more helpful, fair, and useful, less biased, and harder than NCME members expected them to.

Across all of these items, approximately four times as many teachers expressed positive feelings toward standardized tests than was predicted by "experts." Generally less than 10 percent of the teachers expressed negative reactions to tests on these dimensions. Although such percentages are not negligible or unimportant, they are far smaller in most instances than the corresponding percentages of teachers who reacted positively toward tests. Overall, it is again apparent that teachers' general reactions toward tests are neutral to positive.

Teachers were also asked their opinions of four test-related policies. Almost two-thirds of the teachers favored use of competency tests for high school graduation. This compares very closely with the percent of adults in general who favor such test use (Gallup, 1978). Interestingly, high school teachers favored such test use by a greater percent than did either primary or elementary teachers (66 percent versus 53 percent and 63 percent, respectively). About one teacher in four

Table 7. Comparisons Between NCME Members' Impressions and Teachers Who Consider Standardized Achievement Test Results Useful for Various Purposes¹
(Grades K-12)

Standardized Test Results are Useful to:	Teachers' Opinions	NCME Member Impression of Teachers' Opinions
Report to newspapers	10	33
Report to boards of education	52	77
Report to parents	67	62
Report progress to students	56	59
Measure educational status of individuals	61	65
Measure educational "growth" of individuals	77	56
Screen special education students	56	73
Help plan instruction for individuals	63	32
Help plan instruction for class groups	65	47
Detect system-wide general strengths/weaknesses	75	81
Help evaluate teaching procedures or methods	34	15
Help evaluate instructional materials	41	23
Help evaluate teacher performance	21	5
Compare students with national peer groups	58	91
Compare classes in a school	30	36
Compare schools within a system	36	41
Compare a system with systems across the country	56	75

¹Numbers represent percents.

Table 6. Comparisons Between NCME Members' Impressions of Teachers' Personal Feelings About Standardized Achievement Tests¹ and Teachers' Actual Feelings (Grades K-12)

Dimensions ²	Teachers' Feelings	NCME Members' Impressions of Teachers' Feelings
Calm		
1-2	42	10
3-5	51	73
6-7	7	17
Anxious		
Comfortable		
1-2	38	6
3-5	55	80
6-7	7	14
Uncomfortable		
Interested		
1-2	37	16
3-5	55	76
6-7	8	8
Uninterested		
Knowledgable		
1-2	34	9
3-5	64	78
6-7	2	13
Not Knowledgable		
Supportive		
1-2	32	8
3-5	63	83
6-7	5	9
Antagonistic		

¹Numbers represent percents.

²Although the questionnaire used a seven-point scale data were collapsed into a three-point scale range for summary purposes.

Table 9. Teachers' General Opinions of Standardized Achievement Tests and NCME Members' Impressions of Teacher Opinions¹

Dimensions ²	Teachers' Opinions	NCME Members' Impressions of Teacher Opinions
Easy		
1-2	2	0
3-5	76	86
6-7	22	14
Hard		
Helpful		
1-2	24	9
3-5	69	85
6-7	7	6
Harmful		
Unbiased		
1-2	18	7
3-5	69	83
6-7	13	11
Biased		
Useful		
1-2	26	9
3-5	64	81
6-7	10	10
Useless		
Fair		
1-2	25	9
3-5	63	86
6-7	12	5
Unfair		
Valid		
1-2	16	8
3-5	72	79
6-7	12	13
Invalid		

¹Numbers represent percents.

²Although the questionnaire used a seven-point scale data were collapsed into a three-point scale range for summary purposes.

avored increased use of testing for school accountability purposes. Interestingly, only about a third of the respondents favored increased use of criterion-referenced tests. Many teachers were apparently unfamiliar with this concept because many wrote next to this question such comments as "what's that?" or "don't know this term." Almost 20 percent of the teachers left this question blank.

Responses of the teachers anticipated by the NCME sample differed significantly from the actual teacher data. NCME respondents expected far more teachers to favor increased training in test interpretation and increased use of criterion-referenced tests than was true, and they anticipated less favorable reaction to competency tests and use of test results for accountability.

The final set of questions asked teachers whether they personally favored various moratoria on tests. Fewer than 20 percent of the teachers favored a moratorium on all standardized tests or on standardized achievement tests; roughly one-fourth favored such a policy on intelligence tests; and about one teacher in three favored a moratorium on state-mandated achievement tests. NCME members' predictions were close to the teachers' responses concerning tests in general and achievement tests in particular, but significantly overestimated teacher opposition to the other two types of tests.

The responses make it clear that teachers, in general, do not favor moratoria against most testing. These data seem to be directly in conflict with test moratorium resolutions of organizations such as the Council for Exceptional Children and, especially, the National Education Association. While the 20 percent favoring a moratorium on all tests or on just achievement tests is by no means a trivial percentage, it is certainly not an overwhelmingly large percentage either. For comparison purposes, for example, higher percentages of teachers favored increased use of standardized tests for school accountability purposes or to help evaluate teachers' performances than favored moratoria for such instruments.

Summary

"Standardized testing" in our schools has become so immersed in controversy and misunderstanding that what is needed to view the battlefield in perspective is to step away from the mudslinging and data-free generalizations currently being offered both by proponents and detractors of such tests. Tests are not now, nor have they ever been, nor will they ever be, appropriately used as the criterion for performance, as the evaluation device, as the arbiter of good versus less-than-good. Tests can serve a legitimate and useful purpose when they are used along with other information as one part of such decisions. This other information can be objective or subjective, quantitative or qualitative, "norm-referenced" or not. And the best decisions about kids, teachers, programs, or schools will be those in which as much and as many types of information as possible are brought into the process.

In this perspective, tests can often play an important, perhaps even a significant, role in decisionmaking. But they are never appropriately used as stand-alone judges of the quality of anyone or anything. Unfortunately, this perspective -- the only one from which standardized testing can appropriately be assessed -- has been lost by the large number of those who discuss and write about tests from either side of the fence.

We consider our survey data, as summarized above, to provide on the whole a refreshing and encouraging look at the reality of test interpretation in our schools. Kids and teachers and, to some degree at least, administrators (Stetz & Beck, 1978) seem to us generally to have a clear and appropriate view of the role of testing. Students generally consider standardized tests to be "OK" and fair; they want to find out how they scored, but aren't really interested in taking more tests. They generally consider such tests less important and, therefore, less anxiety-provoking and less difficult than teacher-made tests.

Teachers generally find tests to be useful -- somewhat. They personally use the results -- to some degree. They find the uses to which test data are put to be appropriate -- generally. They don't feel there's too much testing in their schools, nor do they want to eliminate standardized tests -- but they don't want to administer more such instruments. They're supportive of tests -- but mostly when used for instruction-related purposes. Teachers find tests most useful for the purposes with which they're primarily concerned: instructional planning for individuals and groups.

Overall, we conclude that both students and teachers generally view standardized tests as one component of the ongoing instructional program, sometimes an important component, sometimes not; sometimes useful, sometimes not. They view test results with both limited respect and healthy skepticism.

We read all of these data as indicating that the only concerned parties in this area who have maintained a clear and balanced perspective throughout the excessive defenses of, and protests concerning, tests are the school staffs and the students. Other groups -- the teacher unions, test publishers, instant-expert journalists, measurement specialists, and professional organizations -- have somehow lost track of the path through this morass, but not real users of the results.

A related conclusion, based on the survey of NCME members, is that measurement specialists -- even those who are members of a theoretically practitioner-oriented organization -- generally do not have a good understanding of the opinions about standardized testing held by the users of such instruments. Going beyond the survey data presented, it is noted that the attitudes predicted of teachers by test specialists are consistently closer to the generalizations made by opponents of testing in the popular media than they are to reality. Test specialists may be uncritically internalizing the claims being made in such articles as being the actual opinions held by teachers and students.

In summary, I would like to plea for a more thoughtful and balanced presentation of the issues in this area -- by both sides -- than we're currently witnessing. A reasoned look at this area of school testing clearly reveals the need for reduced reliance of anecdotal (especially "artificial" anecdotal) and hysterical half-truths and data-independent "evidence" supporting one's biases. I realize well what sells articles in such professional journals as Reader's Digest, Women's Day, and National Elementary Principal, in which we see the bulk of these diatribes, but such yellow journalism is unlikely to make any friends for either side. And we should all recognize by now that this is an era in which all those associated with education need all the friends they can get.

2. TEACHERS AND TESTING: A SURVEY OF KNOWLEDGE AND ATTITUDES

James G. Ward, American Federation of Teachers, AFL-CIO

Since the 1970's, the subject of the testing of students for aptitude, achievement, and ability has been one of the enduring issues for controversy. While almost all involved in education have recognized the many limitations of current testing technology and the many abuses made of student testing, some have called for the severe curtailment or elimination of testing while others have favored reforms in the process. The reaction to problems in testing has been similar to the response of the "Progressive" reformers in the first decade of this century who were genuinely concerned about the corruption in Government and other social ills they saw in American cities. Some were so appalled they favored the elimination of cities and a return to an illusory pastoral innocence. Others saw the reality of the urban milieu and worked for reforms in governance and administration to correct the sicknesses of the cities. History has vindicated the pragmatic reformers and so, it seems, it will also in testing.

In 1973, the American Federation of Teachers (AFT) Executive Council adopted a resolution directing the AFT and its staff to study and disseminate information on the use of intelligence, aptitude, and achievement tests. In a 1976 resolution on testing, the AFT called for the responsible use of tests and implored test developers and publishers to improve the understanding of tests among teachers as well as the population at large. This resolution also called upon test developers to improve obvious deficiencies in tests and listed a number of reforms the AFT favored.

This concern for the proper and appropriate use of high quality tests for students led to further action by the AFT.

In 1978, the AFT applied for and received a 2-year grant from the National Institute of Education to prepare training materials and conduct conferences on improving teachers use of standardized tests. Part of the plan was to survey a representative sample of teachers to ascertain their preparation and knowledge in testing, their assessment of the importance of testing to their teaching tasks, and their attitudes toward various issues in testing.

The development of the survey instrument and the design of the survey sample were done by the Center for the Study of Evaluation at the University of California at Los Angeles.

In the fall of 1979, 800 survey instruments were mailed to a sample chosen systematically from the AFT membership files to be representative of all AFT members in elementary and secondary teaching. Returned and usable survey forms numbered 209 (26.1 percent), and this analysis is based on that total.

Sample Demography

Those who returned completed survey forms were evenly distributed over all grades in elementary and secondary schools. The distribution showed that 46 percent were elementary school (K-6) teachers, 45 percent taught in secondary schools (7-12), 7 percent taught in both, and 3 percent did not state the grade level taught.

Those surveyed were also asked to indicate the socioeconomic status of the community their school serves. The respondents were primarily in communities with middle or low socioeconomic status. High socioeconomic status communities were indicated by 6 percent of the respondents, middle socioeconomic status by 45 percent, low socioeconomic status by 39 percent, and 10 percent did not complete this item.

Respondents were also requested to state whether the community the school serves could be characterized as urban, suburban, or rural. According to the responses, 51 percent taught in urban communities, 34 percent in suburban communities, 9 percent in rural communities, and 7 percent did not state community type.

Education and Training in Testing

Before ascertaining the knowledge and attitudes of the respondents, the survey determined the extent of education and training the teachers had had in testing.

As Table 1 shows, 29 percent of teachers surveyed had taken no college courses devoted exclusively to testing and measurement, and 70 percent had taken no inservice session on the subject. Of the remainder, most (52 percent of the total respondents) had taken one or two college courses in testing, and only 19 percent had taken more than two courses. The record on inservice courses, as shown in Table 1, is even worse.

Those surveyed were also asked to indicate the number of college courses and inservice sessions taken devoted in part to testing and measurement. The results are shown in Table 2. The record here is somewhat better. Only 23 percent had no college courses devoted in part to testing, but 65 percent had no inservice devoted in part to testing.

While no cross-tabulations were done, it appears as though about one teacher in five has had no formal college training in testing and measurement and that after college, only about one in three ever get any inservice training in the area. It is clear that more training for teachers in testing is needed, particularly at the inservice level, to improve the use of tests.

Teacher Expertise in Testing

Nine issues or topics in testing were chosen and teachers in the survey were asked to indicate their assessment of the importance to them of expertise in those areas and were also asked to assess their own expertise in each of those areas. Those results are presented in Tables 3 and 4.

As shown in Table 3, teachers felt that "how to select good tests" and "how to judge the appropriateness of required tests" were the most important areas in which teachers should have expertise, followed by "how standardized tests are developed," "how to evaluate minimum competency testing programs," and "how to interpret norm and percentile scores." Also rated high were "how technical qualities of a test influence what scores mean" and "how to use test results in instructional planning." The only topics rated less than 3 in importance (on a scale of 1 to 4) were "truth in testing legislation" and "what rights and responsibilities are associated with testing."

Table 1. Education and Training Devoted Exclusively to Testing and Measurement

Number of Courses	College Courses Taken		Inservice Sessions Taken	
	Number	Percent	Number	Percent
None*	61	29%	147	70%
1	61	29	21	10
2	48	23	17	8
3 or More	39	19	24	12

*Includes no response

Table 2. Education and Training Devoted In Part to Testing and Measurement

Number of Courses	College Courses Taken		Inservice Sessions Taken	
	Number	Percent	Number	Percent
None*	49	23%	135	65%
1	44	21	27	13
2	38	18	16	8
3 or More	78	37	31	15

*Includes no response

Table 3. Teachers' Ratings of Importance of Expertise
in Areas of Testing

Area of Testing	Average Rating (1=Low, 4=High)	Percentage Indicating	
		High Importance	Low Importance
How to select good tests	3.4	57%	5%
How to judge the appropriateness of required tests	3.4	52	4
How standardized tests are developed	3.3	46	6
How to evaluate minimum competency testing programs	3.2	41	4
How to interpret norm and percentile scores	3.2	40	4
How the technical qualities of a test influence what score means	3.1	37	7
How to use test results in instructional planning	3.1	34	7
Truth-in-testing legislation	2.8	23	12
What rights and responsibilities are associated with testing	2.7	22	15

Table 4. Teachers' Estimated Level of Their Own Expertise
in Areas of Testing

Area of Testing	Average Rating (1=Low, 4=High)	Percentage Indicating	
		High Importance	Low Importance
How to use test results in instructional planning	3.0	27%	7%
How to interpret norm and percentile scores	2.8	24	12
How to judge the appropriate- ness of required tests	2.7	21	11
How to select good tests	2.5	12	15
How the technical qualities of a test influence what score means	2.4	12	20
How to evaluate minimum competency testing programs	2.3	12	22
What rights and responsibil- ities are associated with testing	2.2	13	27
How standardized tests are developed	2.1	7	32
Truth-in-testing legislation	1.9	6	41

As Table 4 shows, only one topic was rated by teachers as high as 3 on the same 1-to-4 scale in their own estimated level of expertise. The area rated at 3.0 was "how to use test results in instructional planning." Other areas ranked high in estimated expertise were "how to interpret norm and percentile scores," "how to judge the appropriateness of required tests," and "how to select good tests." Particularly low levels of expertise were indicated for "how standardized tests are developed" and "truth-in-testing legislation."

The greatest discrepancy exists in the area of "how standardized tests are developed," where importance was rated at 3.3, but expertise at 2.1. Other areas where a great discrepancy was indicated included "how to select good tests," "how to evaluate minimum competency testing programs," and "truth-in-testing legislation."

Use of Standardized Tests

Survey recipients were asked, compared to other information they have about students, to what extent do they use standardized test results in their instructional planning. On a scale from 1 (unimportant) to 4 (important), the uses were rated as follows:

Diagnosis of individual needs	2.8
Student placement/grouping	2.7
Determining class needs	2.5
Judging student progress	2.5
Modification of your course content	2.4
Evaluation of your instructional program	2.4

Teachers indicate greater use of test results for individual student decisions than for group decisions.

Factors That Inhibit Use of Standardized Tests

Conversations with teachers often reveal that teachers feel that standardized tests have problems that inhibit their usefulness to them. Survey respondents were asked to indicate, from a list of eleven problems, which major problems prevent them from more fully using standardized test results. Table 5 presents those results.

The two factors mentioned most often relate to the desire by classroom teachers for more useful information on diagnoses of individual student strengths and weaknesses and prescriptive information to provide guidance for corrective action.

There was also concern with the timing of the return of test results, if they are returned to the teachers at all, the test/curriculum match, and the appropriateness of test materials for some students.

It is clear from these results that teachers use standardized test results to help them make decisions about individual students relating to diagnosis of needs, grouping, and placement, but that teachers recognize the limitations of such tests for these purposes. There seems to be a clearly perceived need for more and better information about students and a desire to improve standardized tests to provide that information.

Table 5. Factors That Inhibit Teacher Use of
Standardized Tests

Factor	Percentage Indicating Factor As A Major Problem
1 Results do not provide an adequate profile of student strengths and weaknesses	63
2 Results do not provide prescriptive information, e.g., guidance as to what materials, instructional activities are needed	52
3 Results are returned too late to be useful, or are not returned to teachers	51
4 Test content does not match my curriculum	47
5 Test materials are inappropriate and/or biased for at least some of my students	46
6 Comparison groups (norms) provided by the tests are not meaningful	29
7 Results are not reported in a form that facilitates interpretation	29
8 Results do not give me any new information about my students	27
9 Tests are given at the wrong time of the year	26
10 Tests take too long to administer	23
11 Technical quality of tests is inadequate	17

Minimum Competency Testing

Slightly below 60 percent of the teachers surveyed stated that their school districts have minimum competency testing programs, and in 81 percent of these cases it is state mandated.

	Number	Percent
No minimum competency testing program	84	40
Has minimum competency testing program	125	60
State mandated	101	48
Locally mandated	24	11

Those in districts with minimum competency testing programs were asked to answer a series of questions about their programs. Those results are presented in Table 6.

The results show the following findings:

- Teachers were not involved in selecting competencies.
- Teachers were not involved in setting passing scores.
- Teachers are mixed as to whether competencies reflect classroom goals.
- Competency test results seem to provide information for planning remedial work, but teachers receive little assistance in doing this.
- Teachers are mixed as to whether school programs have changed much as a result of competency tests, and are also mixed about whether important curricular goals have been ignored.
- Teachers are divided, with a slight leaning toward a positive opinion, about whether competency tests will increase quality instruction.

Given these findings, it is interesting to look at teachers' general attitudes toward minimum competency testing. These results are presented in Table 7. While a large segment of teachers expressed a neutral attitude (34 percent), 37 percent indicated a generally favorable attitude and 20 percent an unfavorable attitude.

Conclusion

Teachers seem to be generally supportive of standardized tests. They understand the need for quality standards in education, and they want more useful information about their students. Particularly, teachers want information to help them in:

- diagnosing individual student needs
- planning instructional activities and programs
- placing and grouping students.

To teachers, testing should be an integral part of the instructional programs, and there should be a high match between tests and curriculum. Teachers seem somewhat supportive of minimum competency testing, but they see many serious flaws in most current programs in this area.

Most significantly, teachers want to improve standardized tests, not eliminate them.

Table 6. Details About Minimum Competency Testing Programs

Questions	Percentage of Respondents Indicating				
	To A Great Extent	Mostly Yes	Mostly No	Not At All	Not Sure/ No Response
a. Do you believe that teachers were involved in selecting the competencies to be tested?	5	11	10	45	29
b. Do you believe that teachers were involved in setting passing scores?	4	9	10	43	34
c. Do the competencies reflect your classroom goals for students?	11	32	22	20	15
d. Do the results provide sufficient information so that remedial work for students can be planned?	17	27	26	8	22
e. Has your school or instructional program changed because of the test?	17	21	24	23	15
f. Have important curricular goals been ignored due to emphasis on minimum competency?	14	20	19	29	17
g. Do you receive assistance in interpreting the test results and/or planning remedial work for students?	9	15	23	39	15
h. Do you think minimum competency testing will help promote higher quality instruction for students?	18	28	18	17	18

Table 7. Teacher Attitudes Toward Minimum Competency Testing

Attitude	Number	Percent
Very Favorable	38	18
Favorable	40	19
Neutral	71	34
Unfavorable	24	12
Very Unfavorable	16	8
No Response	20	10

3. ELEMENTARY TEACHERS AND ACHIEVEMENT TESTS

Leslie Salmon-Cox, Learning Research and Development Center

Several of you have heard reports of this research already. Without further apology for repetition, let me start by making clear what it is that I am here reporting on. The results discussed here focus on suburban elementary school teachers and their uses and perceptions of standardized achievement tests. These data were collected in 1979. At the same time, we worked also in one middle school; in 1980 we interviewed in and studied a senior high school. Also in 1980, we worked in three more elementary schools in a large urban school district. Our interview schedules included questions about teachers' uses and perceptions of IQ tests as well as achievement tests, though I will not discuss this today. Thus, today's discussion centers on suburban elementary teachers and achievement tests.

Organizational and Historical Contexts

With these caveats carefully delineated, let me take one minute to place my research in its more appropriate, larger context. The work I am doing with teachers is part of a three-part program of research entitled "The Social Functions of Educational Testing," funded by the Carnegie Corporation of New York. Under Daniel Resnick's direction, an historical and cross-cultural analysis is being done of the uses of tests in various schooling contexts. Under Lee Sproull's direction, a study, quite parallel to my own, is being done of central district administrators and their uses of and beliefs about standardized testing. Lee has worked in a number of suburban districts, in the same -- nameless -- large urban district we are working in this year, in a parochial system, and in several private schools. In any district in which we have worked with teachers, Lee has also studied the central administration. (Because interviewing central staff in some districts means only one or two relevant people, Lee has worked in far more districts than we.) Finally, our project, my project, is at the school-building level. Our focus is on teachers and other in-school professionals -- principals, guidance counselors, any other relevant actors residing in the school. (In the urban district this includes Title I special teachers, special education teachers, social workers, and others.) That's the organizational context for the work.

The intellectual context is this: several years ago, Lauren and Daniel Resnick, and a small group of us working with them, became interested in the growing controversy that surrounded the use of standardized tests. We spent a background year studying, reading, and discovered that there were almost no recent research-based answers to questions such as the following: What is the actual role for testing within schooling? How is test information actually used? What do various actors, at varying levels of the schooling process, think about standardized tests? What functions are being served by tests and test information within the educational process?

Of course, we all know some of the history. It is ironic, actually, to recall the etiology of standardized testing. Developed in Europe, largely, at the end of the 19th century, standardized mental-abilities testing "caught on" in the United States by World War I, and later development became preeminently an American concern. These tests, we must recall, were designed first, in part, to identify individuals with special qualities; they were designed to find individual students

with talent or deficiencies; they were designed to combat subjective, albeit not always biased, judgment.

There was, of course, always, another reason for standardized tests as well. The archetypical standardized achievement test use, to assess school or teacher effectiveness, dates back in this country to Horace Mann and schooling in Massachusetts in the 1840's and 1850's.

But the heyday for testing came later. Between World War I and 1960, the uses of standardized tests proliferated and became institutionalized. Testing, per se, took on the quality of a social movement (as did the later anti-testing drive). This happened -- speaking generally -- for three reasons. The first, I've mentioned already: the belief that testing was a reliable and valid technology for better understanding individual differences among students. (This, at a time when some eastern urban school districts were faced with anything from 30 to 50 percent enrollments of children in whose homes, at least, English was a second, if existing, language.)

Secondly, school personnel were under pressure to justify their management as rational, productive, efficient, even scientific. As education became a profession, its practitioners became constrained to prove that the quality of the process -- if not the product -- could be measured in "big business" terms. Tests, and their use, bore a stamp of respectability, of "scientificity" -- a hallmark showing careful management, planning, and assessment.

Finally, these two goals, or purposes, for testing were coincident with -- though not accidentally -- the emergence of psychometric technology. As the schools discovered or created needs, the scientific community was developing tools. And the development of these tools was a process enhanced and refined by their use. Or so it was assumed.

There is the rub. We cannot fully know to what extent school professionals of the recent past actually used the information generated by standardized tests. Daniel Resnick's historical analysis of the Pittsburgh Research and Measurement Bureau (Resnick and Schumacker, 1980) does reveal teacher hostility to student examinations -- as selection devices for high school -- and to teacher examinations -- as devices for certification. These two uses for test information were abandoned around World War I. Certainly the recent controversy that has arisen around testing assumes use. (We do know that standardized tests have changed very little in the past 50 years, while schooling, and more important, concepts of schooling, have changed. Perhaps it is this that underlies much current debate.)

It was, then, in the context of the current debate that we began our work. It is important to understand that we began, and largely continued, without strong opinions. We are neither pro- nor anti-testing, at least not along the dimensions that much of that debate is cast in. We view it as our task to produce information, knowledge about actual current practice and reality, on the basis of which informed opinions may be drawn. We have every intention of concluding this work in such a way as to be highly policy relevant. But we most certainly did not begin with foregone conclusions to be proven.

We are, all, of course, familiar with some of these foregone conclusions:

- Test scores label children.
- Standardized tests are more objective than teacher judgment and play an important role in identifying "hidden talent."
- Standardized testing is dictating school curricula, both the content and format of instruction.
- The newer designs for standardized tests provide diagnostic information highly useful to classroom teachers.
- And, at some more vague level, teachers dislike standardized tests and want very much to be rid of them.
- Teachers find standardized tests useful and valuable and would be disturbed if such tests were not available to them.

Overview: Research on Teachers and Tests

Opinions abound about how tests ought to be used, what kinds of tests are appropriate for what purposes, and especially about the effect of the use of test information on individual students. A problem central to the points of controversy is this: whether one advocates continued and more testing because of the benefits thought to accrue, or one proposes test modification or abolition because of the harm thought to ensue -- in either case, a presumption is that test information is being used. And that presumption rests finally on a notion of how teachers behave and think, rests on a conviction that teachers are using the information generated by standardized tests. Our research, to date, indicates that the presumption is not accurate. The teachers with whom we have worked make little use of the information that is available to them from such tests. These teachers do indeed make many and frequent decisions which affect individual students, but standardized test information rarely plays a role in these decisions.

Further, standardized testing, per se, is not a matter of great concern to these teachers. They do make some use of the information; they do have opinions about tests. But overall, such testing is of low salience to the classroom teachers we have studied.

The findings I am about to detail emerge primarily from open-ended but structured interviews, of which two were conducted with each participant. In addition, we have spent time and observed in the schools being studied -- at teacher meetings, at inservice workshops, in the teachers' lounges, and finally, in some classrooms during the administration of spring achievement testing.

A word about data analysis: all of our interviews were tape-recorded, the tapes transcribed as scripts, and the scripts coded by a coding scheme that emerged from the data. All categories of responses, then, which you will see on the tables, were derived from what people said rather than imposed on the responses.

The three elementary schools studied last year offered some interesting contrasts. (This was not entirely serendipitous. I conducted a mailed survey about testing practices in the 46 school districts comprising the surrounding Intermediate

Unit prior to district selection. The responses to that survey were one criterion in our selection procedures.)

If you look at Figure 1, you will see that these three schools varied in:

- their classroom organization and tracking patterns
- the authority pattern of the principal
- the frequency of staff meetings and the presence or absence of an overt school "philosophy"

	<u>School 1</u>	<u>School 2</u>	<u>School 3</u>
Grades in School	K - 4	K - 4	K - 6
Classroom organization	"Self-contained"	"Open" classroom,	"Self-contained"
Tracking pattern	K-2; Departmentalized, 3-4, Heavily tracked	K-4; Team teaching. Moderately tracked.	K-3; Departmentalized, 4-6. No tracking.
Principal; Authority pattern	White, male, Ph.D; Persuasive, by example.	White female, M.Ed.; Friendly authoritarian; highly directive.	White, male, M.Ed.; Non-interacting.
Numbers, sex, classroom teachers	13, 9 F, 4 M	10, 10 F	14, 12 F, 2 M
Numbers interviewed	13	10	12 (10 F, 2 M)
Frequency of formal staff meetings	Weekly	Weekly	Apparently never
Presence/Absence of overt school "Philosophy"	Absent	Present	Absent
Socioeconomic status of community served	Middle to upper SES	Lower middle/upper lower SES	Lower middle/upper lower SES
Attitudes toward community served	Positive; parental interest perceived to be high; teachers assume most parents interested and concerned	Very positive; parental interest perceived to be medium to low; teachers assume most parents caring but unable to know how to make a difference	Somewhat negative; parental interest perceived to be low; teachers assume most parents disinterested.

Figure 1. School Characteristics

- the socioeconomic status of the community served
- the teachers' attitudes towards that community.

Even with these variations in important aspects of the schooling process, what we have found is that -- on the issues in which we were interested -- points of agreement and consensus far outweigh disagreement. Taken together, the data converge to explain how teachers use test information, what they think are appropriate or inappropriate uses for such information, and why, given their general orientations and the necessities of classroom life, test information is of little importance.

Teachers' General Orientations

Though our central focus is on standardized testing, we wanted to be able to place teachers' ideas about that within a broader context. So we began our interviews at a more general level, asking teachers to discuss with us their notions about their roles, the role of schools and schooling, the individual differences among their students, and their general assessment techniques.

When we explored with teachers their major goals, some enumerated a vast number of specifics, others spoke in more general terms. As can be seen from Table 1, all of the teachers mentioned goals that referred to cognitive learning domains as well as to social skills areas.

Table 1. Teachers' Goals
(by teacher; N = 35)

	Totals
Number who mentioned cognitive goals	35
Number who emphasized	10* (29%)
Number who mentioned social goals	35
Number who emphasized	20* (57%)

*Not all responses were codable for emphasis.

Goals in the latter area -- social skills; working with children to better their discipline, their behavior, their self-concepts, their ability to get along with others -- were given far more emphasis by most than goals in the cognitive domain.

In response to our question, "What do you think is the role of schools and schooling in American society?" we see in Table 2 (which, you will note is coded by response since multiple responses were possible) again an emphasis on the socialization functions of schools, though basic skills instruction is also frequently mentioned.

Table 2. Perceptions of the Role of Schools and Schooling
(by response; N = 40)

Number of responses mentioning basic skills instruction	16 (40%)
Number of responses mentioning "socialization," "dealing with the whole child"	24 (60%) 40 (100%)

Before leaving the subject of goals and the role of schools, I'd like to add that while these teachers attached great importance to social goals and functions, they did not do so in lieu of basic instructional goals. Rather, from our conversations with them, it was clear that they consider such instruction to be the taken-for-granted responsibility of the school. In addition, then, they felt a need to emphasize other schooling functions as well.

Table 3 shows responses to a question about what were considered to be the important individual differences among children that classroom teachers ought to be aware of.

Table 3. Significant Individual Differences Among Students
(multiple responses possible; N = 35)

	Percentage of teachers responding in each category
Family/home background	63
Social skills/maturity	31
Ability	43
Special education needs	9

Again, multiple responses were possible. Of the teachers responding, 63 percent of them pointed to family and home background differences as those which most significantly distinguished among their students. While "ability" differences were the second most frequently mentioned, fewer than half of the teachers cited differences in ability as the significant ones.

Tables 4 and 5 tell us something of teachers' general assessment techniques. (In Table 4, we have indicated the range of responses across the 3 schools studied because there were some marked variations school by school on this question.) Overwhelmingly, we see that the most frequently mentioned form of assessment was observation; in effect, "I watch my kids, I see what they're doing -- as they work

Table 4. Teachers' General Assessment Techniques
(multiple responses possible; N = 35)

	Percentage of teachers responding in each category
Observation	80% Range: 75% - 85%
Interaction	49% Range: 30% - 67%
Daily paperwork	46% Range: 31% - 58%
Teacher-made tests	34% Range: 8% - 75%

*Of the total 35, 17 teachers (49%) rated "observation" as the single most important technique.

Table 5. Timing of General Assessment
(by teacher; N = 29*)

	Percentage of teachers responding in each category
Immediately; 1st week or 2	17
Within first month; 6 to 9 weeks	52
By mid-year; by Christmas	28
Later; some point beyond mid-year	3

*Six respondents did not answer this question.

alone, with other children, in the classroom, on the playground, and so on." The other techniques mentioned are arrayed in Table 4. As noted there, 49 percent of the teachers emphasized observation as the single most important technique they used to assess progress. None of these teachers mentioned standardized tests in this context.

A related question was one on timing, when do teachers "get a feel" for their students, for what their students will be able to accomplish. We see in Table 5 that more than two-thirds of them said that they came to some general assessment of their students fairly early in the school year.

The following statements sum up these teachers' general orientations:

- While each has cognitive, instructional goals for students, emphasis is placed on social goals.
- There is a strong belief that the role of the school is a socializing one, its function is to "deal with the whole child."
- What each child brings to school from his or her home background is what most significantly differentiates students. While ability differences are salient, they are not, of themselves, the most salient in the classroom.
- Reliance on observation of children is the most frequently used and a most highly trusted method of assessing children's progress in school.
- Judgments about a child's performance, or possible performance over the school year, are made fairly early on in that school year.

When we think of the characteristics of standardized tests as performance measures for students and contrast these with the general orientations sketched above, it is not difficult to begin to see why standardized testing is not of great relevance to the classroom teacher. Let's look at just how these tests are used and viewed.

We see in Table 6 the responses to our question about actual achievement test information use. Responses grouped themselves in five categories. Teachers could mention as many or as few uses as they wished, and there were 98 responses altogether.

Table 6. Achievement Test - Use
(multiple responses possible; by response)

		School 1.	School 2	School 3	Totals
As reflection on/guide to/instruction	Total	8 (26%)	4 (14%)	5 (13%)	17 (17%)
As supplement to/confirmation of/information on students	Total	11 (35%)	12 (41%)	19 (50%)	42 (43%)
Placement/tracking/grouping	Total	9 (29%)	6 (21%)	13 (34%)	28 (29%)
Communication	Total	1 (3%)	5 (17%)		6 (6%)
Miscellaneous	Total	2 (6%)	2 (7%)	1 (3%)	5 (5%)
		31 (100%)	29 (100%)	38 (100%)	98 (100%)
Average number of responses per teacher		2.4	2.9	3.2	

Using test information as a device with which to reflect upon instruction or curricula was one cited use. These teachers did not however, in the main, view the information in this way. Rather, their major use of the information was as a supplement to, or confirmation of, other kinds of information they already had about children. Especially important was the confirming function of test information: feeling strongly that they know their students through their observation and work with them, the teachers think of the tests as yet one more piece of evidence, one further measure, of how a child is doing in school.

Achievement test information also plays a role in grouping and tracking students. The role appears to be overwhelmingly again a confirmatory or supplementary one. This is particularly interesting, as the "selection" function of standardized tests is much discussed. While the tracking pattern varied in the schools studied, in none of the three schools was testing observed to be of much consequence vis-a-vis tracking.

With regard to grouping (within class instructional grouping during the year), test scores are cited as playing some role. This was most often the case in School 3. In this "untracked" school, there was random assignment from grade to grade, and very little information was passed on with the children. It is very likely, then, that past test scores play a larger role here than in Schools 1 and 2, which are "tracked," and where information is richer.

"Communication" uses refer to conversations with parents in Table 6, though there were one or two mentions of school board or community-in-general when "appropriate" use was asked.

Tables 7 and 8 display teacher responses to questions about their notions regarding the ways in which achievement test information should be used (as distinct from how it is used).

Table 7. Achievement Test - Appropriate Use
(multiple responses possible; by response)

		School 1	School 2	School 3	Totals
As reflection on/guide to/instruction	Total	4 (21%)	2 (11%)	4 (21%)	10 (18%)
As supplement to/confirmation of/information on students	Total	6 (32%)	9 (47%)	6 (32%)	21 (37%)
Placement/tracking/grouping	Total	4 (21%)	4 (21%)	5 (26%)	13 (23%)
Communication	Total	2 (11%)	2 (11%)	4 (21%)	8 (14%)
Miscellaneous	Total	3 (16%)	2 (11%)		5 (9%)
		19 (100%)	19 (100%)	19 (100%)	57 (100%)
Average number of responses per teacher		1.5	1.9	1.6	

Table 8. Achievement Test - Inappropriate Use
(multiple responses possible; by response)

		School 1	School 2	School 3	Totals
Those relating to instruction/ curricula	Total		1 (8%)	2 (18%)	3 (8%)
Those relating to judgments about students	Total	4 (33%)	2 (15%)	2 (18%)	8 (22%)
Those relating to decision- making	Total	5 (42%)	6 (46%)	2 (18%)	13 (36%)
Miscellaneous	Total	3 (25%)	4 (31%)	5 (45%)	12 (33%)
		12 (100%)	13 (100%)	11 (100%)	36 (100%)
Average number of responses per teacher		.92	1.3	.92	

It is predictable that Table 7 reports responses quite similar to those in Table 6. Unless under severe role conflict, people generally consider their actual behavior to be appropriate. Among our teachers this was so clearly the case that the percentage of responses in several categories is quite similar in Tables 6 and 7. It is interesting that teachers in School 3 had fewer responses regarding tracking and grouping as appropriate than was the case when they were asked about actual practice.

Table 8 shows teachers' ideas about inappropriate uses of achievement test information. At first glance, these data were surprising. That is, it could have been the case that teachers would find inappropriate those very uses for test information which were least common for themselves. (In that case, Table 8 might have been a reversal of Table 6.) In fact, this is not what the data tell us. Rather, teachers identified the following as inappropriate:

- (1) exaggerations of the uses they already engaged in, and,
- (2) the use of achievement test information for decisionmaking in isolation, that is, without regard to other information as well.

An interesting aspect of the data in Table 8 is the decline in response numbers. This is true for Tables 6, 7, and 8, as number of responses per teacher declined with each question.

Misuse or inappropriate use was a question requiring ample probing. Among this group of teachers, clearly this is not a matter of high salience.

Finally, at the last moment, I decided to include one more set of data that combines findings from Lee Sproull's administrative study with the work we have done with teachers. In each case, we have asked our respondents, "If standardized

testing were abolished, who would care? Who would miss it?" In addition, Lee asked her group of administrators, "Who finds test information most useful?" Looking at Table 9 (Sproull and Zubrow, 1980), we see that central office staff perceive actors at the building level to be those for whom the information is most useful. This is true across the four settings in which Lee worked. And in response to the second question -- who would miss testing -- again we see attributions to the building level, especially to teachers.

Now, if we turn to the teachers themselves, in Table 10 we see another kind of attribution, to parents. In fact, only 16 percent of the responses were of the order, "I would miss them." (Note that teachers' responses are broken down to indicate responses to self, vs. statements such as "I wouldn't, but maybe other teachers would.") Teachers assume that parents want test information and would feel its loss. The "parent" group is further subdivided because some teachers made a careful distinction between "parents who care" (and these were frequently identified as parents of children doing well in school) vs. all parents.

With regard to standardized testing, everybody assumes that there is someone who wants it, uses it, but few claim personally to be its consumers.

In conclusion, among these teachers, the lack of salience we found regarding tests and test information is crucial to the debate that rages around testing. Clearly, if test information is largely unused, it can be neither harming nor helping in the ways in which its critics and advocates suppose.

While it is claimed that tests are determining curricula, in these schools teachers rarely use test information to ponder their instruction or curricular content. (We do expect, however, that we will find differences on this dimension as we move "up" the schooling ladder. When teachers become more content-oriented rather than child-oriented, test information may be more influential in shaping instruction.)

There is concern that test scores "label" children, providing information which prejudices their possible achievement. Our data suggest that teachers share this concern and are leery of using test scores in this manner.

Some claim that achievement tests can help to diagnose student strengths and weaknesses and are valuable to teachers for this purpose. While we discovered a great "hunger" for diagnostic instruments among the teachers with whom we spoke, we did not find achievement tests much used in this manner. Teachers desire diagnostic tests that are precise, highly matched to curricula and instruction, and the results of which are delivered and can be used, in timely fashion. Achievement tests do not much match these criteria.

What is overlooked too frequently is the teachers' reality. Of our group of elementary teachers, the following facts are real:

Teachers make decisions constantly, some long-range, some short-range, many each day. They do so primarily by observing and interacting with their students in a variety of contexts. They are comfortable with their techniques for assessing students. And, these techniques rarely include information generated by standardized achievement tests.

Table 9. Salience of Test Information to Various
Actors as Perceived by Central Office Actors
(multiple responses possible)

Salience Indicator	School System Type				Total
	Urban	Suburban	Parochial	Private	
Who finds it Most Useful?					
<u>Central Office</u>					
Superintendent	-	1(4%)	-	-	1(2%)
General Administrator	-	4(16%)	-	3(38%)	7(14%)
Instructional Administrator	3(19%)	4(16%)	1(50%)	-	8(16%)
Pupil Personnel Administrator	-	1(4%)	-	-	1(2%)
<u>Buildings</u>					
Principals	2(13%)	6(24%)	-	-	8(16%)
Teachers	9(56%)	7(28%)	1(50%)	3(38%)	20(41%)
Counselors	1(6%)	2(8%)	-	2(25%)	5(10%)
Other	1(6%)	1(4%)	-	-	2(4%)
responses n =	16	25	2	8	51
Who Would Miss Testing the Most If It Were Abolished?					
<u>Central Office</u>					
Board	1(4%)	3(7%)	-	-	4(5%)
Superintendent	-	1(3%)	-	-	1(1%)
General Administrator	6(21%)	3(7%)	-	1(20%)	10(13%)
Instructional Administrator	2(7%)	3(7%)	1(50%)	-	6(8%)
Pupil Personnel Administrator	-	-	-	-	0(0%)
<u>Building</u>					
Principals	1(4%)	-	-	-	1(1%)
Teachers	8(29%)	11(29%)	-	1(20%)	20(27%)
Counselors	2(7%)	8(20%)	1(50%)	-	11(15%)
Special Education and Itinerant Staff	1(4%)	3(7%)	-	-	4(5%)
Parents, students, and community	4(14%)	6(15%)	-	3(60%)	13(17%)
Other	3(11%)	2(5%)	-	-	5(7%)
responses n =	28	40	2	5	75

Note: Percentages may not sum to 100 due to rounding.

*From "Standardized Testing in the Educational Organization Administrative Performance Information System" AFRA, 1980.

Table 10. If Standardized Tests Were Abolished, Who'd Care?
(multiple responses possible; by response)

	Elementary
No one	2(4%)
Teachers	13(26%)*
Self 8(16%)	
Other 5(10%)	
Parents of some	9(18%)
Parents	9(18%)
School administrators	8(16%)
School Board	5(10%)
Guidance counselors	4(8%)
	50

*Of group citing teachers in elementary schools (13), 8 report using achievement tests, 3 basal tests, 1 readiness test.

Note: Further research with teachers in urban schools, at all grade levels, confirmed these initial findings. Current research is being conducted to ascertain teachers' attitudes towards, and uses of, objectives-based instructional programs. A key component of these programs is frequent diagnostic testing of the kind teachers have long claimed to want and need.

4. A SURVEY OF THE USE OF VARIOUS ACHIEVEMENT TESTS

Jennie Yeh, Center for the Study of Evaluation
Joan Herman, Center for the Study of Evaluation
Lawrence M. Rudner, National Institute of Education

Testing is a fact of life in American public schools: On the average, a child takes six full batteries of standardized achievement tests before graduating high school (Houts, 1975). Two hundred million achievement test forms are used each year in the United States, and this number represents only about 65 percent of all educational and psychological testing that is carried out. According to the Government Accounting Office, 90 percent of the local education agencies throughout the country administer standardized norm-referenced tests to children within their purview (HRD-76-113). Further, a recent report estimated that the cost of administering standardized tests to elementary school students in 1976 was well over a quarter of a billion dollars (EDC News, 1977).

While these figures in themselves suggest the magnitude of the testing enterprise, they do not include the other kinds of measures that are administered routinely in classrooms -- criterion-referenced tests, curriculum-embedded tests, and teacher-developed tests. There are no accurate figures reflecting the prevalence of these other assessments, but there is little doubt that their use is more widespread in schools than standardized achievement measures.

American schools, then, certainly devote considerable time and financial resources to testing students. Despite these significant expenditures, however, we know very little about how the results of these efforts are used in schools. What functions do tests serve in classrooms? Of what use are test results? Studies a decade ago reported little use (Goslin, 1965, 1967; Skager, 1969), and more recent studies of standardized tests similarly find scant evidence of teacher use (Boyd et al., 1975; Airasian, 1980; Beck and Stetz, 1979). Have newer forms of assessment, e.g., the emergence of criterion-referenced and curriculum-embedded tests, altered the picture?

The study reported here provides some preliminary answers about teachers and testing. In order to develop more descriptive knowledge about the extent of testing and the use of tests in school, the study surveyed teachers about:

- (1) the frequency and diversity of various forms of reading and mathematics measures used in schools;
- (2) the use of these measures for instructional decisionmaking;
- (3) contextual variables surrounding test administration and interpretation;
- (4) teacher knowledge and background in testing.

In addition to providing basic descriptive information, the study also sought to identify factors that influence test use.

The Study

Sample. Previous studies (Goslin, 1965, 1967) indicated that teachers' testing practices varied with the location of schools and the socioeconomic status (SES) of parents. Consequently, twenty schools covering a wide range of locations and SES groups were sampled. Five California districts -- representing urban, rural, and suburban schools -- were included. In each of these districts, four schools were purposively sampled on the basis of SES and receipt of outside funding (e.g., ESEA Title I).

All kindergarten through sixth-grade teachers at each school were sent questionnaires. Two hundred and sixty completed questionnaires were received and processed. Extrapolating from the previous year's public school directory, the return rate was estimated to be 60 percent.

Instrument. A questionnaire was developed for teachers to gather information related to five areas: teacher background characteristics; classroom organization and typical instructional practices; methods used to assess progress; use and understanding of required tests; opinions of, and attitude toward, current testing practices.

Analysis. While raw frequency counts, percentages, and means were computed for each question, additional information was obtained by regrouping the respondents by demographic characteristics. Respondents were classified by amount of teaching experience, classroom organization, grade level taught, and whether they had paid aides. Multiple one-way analyses of variance were computed by each classification on each appropriate question. It was recognized that the large amount of significance testing would yield occasional spurious results with regard to group differences.

Findings with Regard to Frequency, Diversity, and Purpose. Sources of information used by teachers for initial reading and mathematics placement are presented in Table 1. Teachers most frequently used some form of test results to assess student skills in reading and math at the beginning of the school year. Fifty-eight

Table 1. Sources of Information Teachers Use to Place Students
In Reading and Math at the Beginning of the Year

Source	Percent of Teachers Reporting Use For	
	Reading Placement	Math Placement
Test results (e.g., standardized, teacher made)	58.7 (148)	65.8 (160)
Previous placement	15.1 (40)	16.0 (39)
Informal assessment (e.g., teacher observation, work sample)	24.2 (61)	16.9 (41)
Other	1.2 (3)	1.2 (3)

percent reported using test results most often for initial reading placement and 65.8 percent reported using test results most often for initial math placement. The findings do not show a significant difference between the sources of information used by teachers in self-contained classrooms as compared to team teachers. Test results seem to provide the teacher with a quick and acceptable estimate of ability of those students unfamiliar to the teacher.

Sources of information used by teachers for assessment throughout the year are presented in Table 2. Teachers most frequently used information from interactions with or observations of students, informal assessment techniques (e.g., oral quizzes, reading aloud) or results from teacher-developed tests to assess their students' progress. The least frequently used sources of information were the results from standardized and instructional program or curriculum-embedded tests, while moderate use was made of information derived from students' place in their textbooks and work assignments. Apparently, once initial placements are made and teachers become more familiar with their students' abilities, they are less likely to rely on standardized or instructional program tests to give them information about students' progress.

Teachers' assessment practices occasionally differed significantly with teaching experience and with the presence of an aide. Teachers with aides were more likely to use the results from instructional program tests, students' place in a textbook, and the results from informal assessment techniques. These results suggest that teachers with aides use more assessment methods. With additional help in the classroom, the teacher perhaps has more time for personal involvement with students and also more time to utilize the various sources of information available.

As shown in Table 2, teaching experience was related to differences in assessment practices. Less experienced teachers (1-7 years of experience) were more likely than more experienced teachers (8 + years of experience) to report the use of results from their own tests, student work assignments, and informal assessment, but were less likely to use results from standardized or curriculum-embedded tests. New teachers are apparently as likely as experienced teachers to use intuitive methods (interaction with students, informal assessment) to assess student progress, and are more likely to also use concrete information (their own tests, work assignments, informal assessment).

The frequency with which teachers administer math and reading tests is reported in Table 3. Fifty-nine percent of the teachers reported giving either monthly or weekly reading tests while 8.7 percent reported giving daily reading tests. In comparison, 18.1 percent of the teachers reported giving monthly math tests, 50 percent reported giving weekly math tests, and 9.7 percent reported giving daily math tests. Thirty-two and three-tenths percent of the teachers gave less than 1-2 reading tests a semester, whereas 21.8 percent of the teachers gave less than 1-2 math tests a semester. Clearly, tests are more frequent in mathematics.

The results for both subject areas showed a trend for upper-grade teachers to give tests more often than primary-grade teachers. These results seem to confirm information from exploratory interviews that teachers gave fewer tests to younger children because they felt these students lacked the powers of concentration and skills necessary for taking tests. When primary-grade teachers did give instructional program tests, they used the results to assess student progress throughout the year.

Table 2. Sources of Information Teachers Use to Assess Student Progress Throughout the Year

Frequency of Reported Use
(1=never use, 5=always use)

Context	Source of Information						
	A	B	C	D	E	F	G
	Stand- ardized Test Results	Instruc- tional Program Test Results	Teacher- Developed Test Results	Stu- dent's Place in Book	Work Assign- ments	Inter- action With Student	Informal Assess- ment
Any Paid Aides							
no paid aides	2.3(69)	*2.8(67)	3.9(71)	Δ 3.5(72)	3.8(72)	4.2(72)	*3.8(71)
some paid aides	2.2(151)	3.2(146)	3.9(153)	3.8(150)	3.7(148)	4.3(154)	4.1(151)
Teaching Experience							
1 - 7 years	2.1(63)	2.9(60)	*4.0(64)	Δ 3.9(64)	*3.9(63)	4.4(64)	Δ 4.2(64)
8 - 15 years	2.4(91)	3.3(90)	3.7(92)	3.6(92)	3.7(90)	4.2(93)	4.0(91)
$\geq$ 16 years	2.4(86)	3.1(83)	3.9(88)	3.6(86)	3.5(86)	4.2(89)	3.8(81)
Class Organization							
team-taught classroom	2.3(19)	3.3(77)	3.9(81)	3.7(79)	3.6(78)	4.2(81)	4.0(79)
self-contained classroom	2.2(134)	3.0(129)	3.8(135)	3.7(135)	3.8(134)	4.3(137)	4.0(136)
Grade Levels Taught							
K - 2nd	2.1(95)	3.2(96)	3.8(98)	3.6(95)	3.6(95)	4.2(98)	3.9(96)
3rd - 4th	2.3(48)	3.1(45)	4.0(50)	3.9(50)	3.8(49)	4.3(50)	4.0(50)
5th - 6th	2.4(58)	3.0(54)	3.9(57)	3.7(58)	3.9(57)	4.3(58)	4.1(58)
TOTAL SAMPLE	2.2(221)	3.1(215)	4.0(226)	3.7(224)	3.7(222)	4.2(227)	4.0(225)

* = one-way ANOVA significant at $\alpha \leq .05$

Δ = one-way ANOVA significant at $.05 < \alpha \leq .10$

Number responding are in parentheses.

Table 3. Frequency of Teacher Test Administration

Percent of Teachers Reporting

	Never	1-2 times a year	1-2 times a semester	Monthly	Weekly	Daily	Median	N
Reading Tests	4.5	111.2	17.0	29.5	29.5	8.7	3.1	213
Math Tests	3.8	8.8	9.2	18.1	50.0	9.7	3.7	210

The types of tests administered by teachers are reported in Table 4. Instructional program and teacher-developed tests constituted the greatest proportion of all tests teachers administered to their students. This finding reinforces the previous finding that results from teacher-developed tests were most frequently used to assess student progress throughout the year. As expected, locally developed

Table 4. Teachers' Use of Different Types of Tests

Proportion of all Tests Given Accounted for by Each Type
(1=0%, 6=100%)

Context	A Standardized Tests	B Instructional Program Tests	C Locally Developed Tests	D Teacher- Developed Tests
Any Paid Aides				
no paid aides	2.3(63)	3.6(68)	*2.2(62)	3.5(69)
some paid aides	2.1(134)	3.6(143)	2.7(127)	3.5(134)
Teaching Experience				
1 - 7 years	2.2(59)	3.6(60)	2.6(58)	3.5(61)
8 - 15 years	2.4(78)	3.6(84)	2.6(75)	3.5(78)
≥ 16 years	2.2(78)	3.5(83)	2.5(69)	3.5(80)
Classroom Organization				
team-taught classroom	2.3(71)	3.8(77)	*2.9(71)	3.6(125)
self-contained classroom	2.1(119)	3.6(127)	2.4(111)	3.6(69)
Grade Levels Taught				
K - 2nd grade	2.1(84)	*3.7(92)	2.7(82)	*3.2(82)
3rd - 4th grade	2.2(42)	4.1(47)	2.5(41)	3.7(48)
5th - 6th grade	2.2(54)	3.3(56)	2.5(49)	3.9(54)
TOTAL SAMPLE	2.2(197)	3.6(211)	2.6(188)	3.5(203)

*One-way analysis of variance significant at $\alpha \leq .05$
Number responding are in parentheses.

tests accounted for a more moderate proportion, while standardized tests constituted the smallest proportion of tests teachers administered. Teachers with paid aides and teachers in team-taught classrooms tended to administer more local tests than teachers without aides and teachers in self-contained classrooms. This is not surprising in light of the time required for test development.

Teachers of different grade levels differed significantly in their administration of instructional program tests. Third- and fourth-grade teachers administered the greatest proportion of this type of test, whereas fifth- and sixth-grade teachers administered the lowest proportion. This finding is consistent with third- and fourth-grade teachers reporting comparatively greater use of instructional program tests results.

The percentage of teachers who develop their own tests is shown in Table 5. Fifty-three percent of the teachers responded that they did develop their own tests, and 47 percent indicated that they did not. Lower-grade teachers were significantly less likely to develop their own tests than were upper-grade teachers. This finding is consistent with lower-grade teachers' general tendency to give fewer tests and their slight tendency to use instructional program tests more frequently.

Table 5. Do Teachers Usually Develop Their Own Reading/Math Tests?

	No - do not develop tests	Yes - develop tests
Any Paid Aides		
no paid aides	40% (26)	60% (39)
some paid aides	50 (73)	50 (74)
Teaching Experience		
1 - 7 years	51 (32)	49 (31)
8 - 15 years	48 (42)	52 (45)
≥ 16 years	48 (38)	52 (41)
Class Organization		
team-taught classroom	49 (38)	51 (39)
self-contained classroom	47 (61)	53 (68)
Grade Level Taught		
K - 2nd grade	63 (59)	*37 (35)
3rd - 4th grade	48 (23)	52 (25)
5th - 6th grade	30 (16)	70 (38)
TOTAL SAMPLE	47% (100)	53% (112)

*One-way analysis of variance significant at $\alpha \leq .05$.

Number responding are in parentheses.

Findings with Regard to Test Qualities Teachers Consider Important. The reasons given by teachers for developing their own tests are shown in Table 6. Teachers reported that the most important reasons for developing their own tests are that these tests more accurately assess the effects of their instruction and are more suitable for their students. This finding confirms the reasons teachers stated for developing their own tests in preliminary study interviews.

Upper-grade teachers were more likely to report that insufficient funds and lack of information about prepared tests were important reasons for developing their own tests. There was no difference in responses reported by teachers with and without aides, and teachers with little or lengthy experience.

Table 6. Why Teachers Develop Their Own Tests

Rated Importance of Various Reasons
(1=very unimportant, 5=very important)

Context	Reason				
	A Insufficient Funds to Buy Tests	B Teacher- Developed Test More Suitable For Students	C No Time To Order Published Tests	D Information About Prepared Tests Unavailable	E Teacher- Developed Tests Better Assess Instruction
Any Paid Aides					
no paid aides	1.9(41)	4.5(42)	1.6(39)	1.9(37)	4.6(41)
some paid aides	1.7(64)	4.5(71)	1.6(65)	1.9(66)	4.6(72)
Teaching Experience					
1 - 7 years	1.6(31)	4.6(33)	1.5(31)	1.8(31)	4.7(33)
8 - 15 years	1.9(44)	4.5(45)	1.5(44)	1.9(44)	4.6(43)
≥ 16 years	1.8(34)	4.5(42)	1.7(33)	1.8(33)	4.8(42)
Class Organization					
team-taught classroom	1.7(34)	4.4(33)	1.5(35)	1.7(35)	4.5(41)
self-contained classroom	1.9(68)	4.6(73)	1.7(66)	2.0(66)	4.6(70)
Grade Levels Taught					
K - 2nd grade	*1.4(32)	4.4(36)	1.6(33)	*1.6(33)	4.5(37)
3rd - 4th grade	2.1(22)	4.7(24)	1.7(21)	2.5(21)	4.7(23)
5th - 6th grade	2.1(37)	4.5(39)	1.8(36)	2.1(36)	4.5(39)
TOTAL SAMPLE	1.8(109)	4.5(113)	1.6(104)	1.9(103)	4.6(113)

*one-way analysis of variance significant at $\alpha < .05$.

Number responding are in parentheses.

The reasons given by teachers for not developing their own tests are shown in Table 7. Teachers reported that the most important factor was that suitable published tests were available. Teachers with aides were significantly more likely to report such availability. There were no significant differences between teachers in self-contained vs. team-taught classrooms.

Table 7. Why Teachers Do Not Develop Their Own Tests

Rated Importance of Various Reasons
(1=very unimportant, 5=very important)

Context	Reason				
	A Comparison of Students to National/ Local Norms Important	B Suitable Published Tests Unavailable	C No Time to Develop Tests	D Testing Only Done When Required	E More Teacher Training In Test Construction Necessary
Any Paid Aides					
no paid aides	2.8(44)	Δ 3.4(45)	2.4(41)	2.5(42)	2.5(41)
some paid aides	2.7(86)	3.8(89)	2.4(83)	2.1(79)	2.4(80)
Teaching Experience					
1 - 7 years	2.8(40)	3.5(40)	*2.8(41)	2.2(40)	2.2(41)
8 - 15 years	2.8(54)	3.8(56)	2.3(53)	2.1(49)	2.8(51)
$\geq$ 16 years	2.6(49)	3.8(51)	2.1(43)	2.6(45)	2.3(43)
Class Organization					
team-taught classroom	2.7(47)	3.7(49)	2.2(45)	2.3(44)	2.3(43)
self-contained classroom	2.9(83)	3.7(86)	2.6(80)	2.3(77)	2.6(78)
Grade Levels Taught					
K - 2nd grade	2.6(64)	3.8(66)	2.3(62)	2.4(57)	2.5(60)
3rd - 4th grade	2.8(28)	3.7(28)	2.8(26)	2.3(27)	2.6(25)
5th - 6th grade	3.1(31)	3.5(32)	2.2(29)	1.9(29)	2.4(29)
TOTAL SAMPLE	2.7(130)	3.7(134)	2.4(124)	2.3(121)	2.5(121)

*One-way analysis of variance significant at $\alpha \leq .05$.

Δ One-way analysis of variance significant at $.05 < \alpha \leq .10$.

Number responding are in parentheses.

Table 8 shows qualities considered by teachers in selecting prepared tests. Clear format, similarity to class material, and accurate prediction of achievement were the qualities teachers considered most important when choosing prepared tests. These qualities were the same as those given by teachers for developing their own tests. Consistent with earlier results, primary-grade teachers were particularly concerned with test format. The quality teachers considered least important was the test publisher's good reputation.

Table 8. Qualities That Teachers Consider When Selecting Tests

Rated Importance of Various Reasons
(1=very unimportant, 5=very important)

Context	Quality				
	A	B	C	D	E
	Test Material Similar to That Presented in Class	Test Simple to Administer and/or Score	Test Has Clear Format, Directions	Test Publisher Has a Good Reputation	Test Accurately Predicts Student Achievement
Any Paid Aides					
no paid aides	4.4(70)	4.2(69)	4.5(70)	2.6(68)	4.4(70)
some paid aides	4.6(149)	4.2(150)	4.7(150)	2.9(146)	4.4(149)
Teaching Experience					
1 - 7 years	4.4(65)	4.1(65)	4.5(65)	2.7(64)	4.4(65)
8 - 15 years	4.6(89)	4.2(90)	4.7(90)	2.7(88)	4.4(91)
≥ 16 years	4.5(84)	4.2(83)	4.6(84)	2.9(79)	4.5(82)
Class Organization					
team-taught classroom	4.6(80)	4.2(79)	4.6(79)	2.7(76)	4.3(79)
self-contained classroom	4.5(140)	4.3(141)	4.6(142)	2.8(138)	4.4(141)
Grade Levels Taught					
K - 2nd grade	4.6(96)	4.3(96)	*4.7(95)	2.8(91)	4.5(94)
3rd - 4th grade	4.7(50)	4.2(50)	4.6(50)	2.9(48)	4.5(50)
5th - 6th grade	4.3(56)	4.0(56)	4.4(57)	2.9(56)	4.3(56)
TOTAL SAMPLE	4.5(219)	4.2(219)	4.6(220)	2.8(214)	4.4(219)

*One-way analysis of variance significant at $\alpha \leq .05$.
Number responding are in parentheses.

Findings with Regard to Ways Test Results are Used. Table 9 displays ways teacher-developed test results are used. Teachers indicated that they used results from their own tests frequently or always for all of the listed purposes (for instructional decisions, evaluation of program effectiveness, grade assignment, information to others). Occasionally responses differed, depending on the presence or absence of aides, grade level taught, and classroom organization. Teachers without aides tended to report using test results for grading more often than did teachers with aides, and lower-grade teachers were less likely to use test results to assign grades than were upper-grade teachers.

Table 9. How Teachers Use Information From Their Own Tests

Frequency of Use for Different Purposes
(1=never, 4=always)

Context	Purpose			
	A To Make Instructional Decisions	B To Evaluate Effectiveness of Methods/ Materials	C To Assign Grades	D To Provide Information To Parents/ Staff
Any Paid Aides				
no paid aides	3.2(68)	3.2(68)	Δ 2.8(67)	3.1(67)
some paid aides	3.2(147)	3.1(147)	2.5(145)	2.9(146)
Teaching Experience				
1 - 7 years	3.3(64)	3.2(64)	2.5(63)	3.1(63)
8 - 15 years	3.1(89)	2.9(89)	2.5(87)	2.9(89)
$\geq$ 16 years	3.2(82)	3.1(83)	2.6(80)	2.9(82)
Class Organization				
team-taught				
classroom	3.3(78)	3.1(79)	2.7(77)	*3.1(75)
self-contained				
classroom	3.2(138)	3.1(137)	2.5(136)	2.9(138)
Grade Levels Taught				
K - 2nd grade	3.2(90)	3.0(91)	*2.4(88)	2.9(89)
3rd - 4th grade	3.2(50)	3.2(50)	3.0(49)	3.0(50)
5th - 6th grade	3.2(55)	3.2(55)	2.8(56)	3.1(55)
TOTAL SAMPLE	3.2(215)	3.1(215)	2.6 (212)	3.0(213)

*One-way analysis of variance significant at $\alpha < .05$.

Δ One-way analysis of variance significant at $.05 < \alpha \leq .10$.

Number responding are in parentheses.

Teachers' responses to students' test failures are shown in Table 10. The most frequent action that teachers reported taking when students failed a test was to reteach the same concept or skill. The least frequent teacher action was allowing the student to progress to the next unit. Responses differed significantly between teachers who had aides and those who did not. Teachers without aides were less likely to reteach, provide in-class tutoring, or assign additional practice with different materials, and more likely to let a student progress to the next unit.

Table 10. What Teachers Do When Students Fail a Test

Relative Frequency of Various Actions
(1=most frequent, 6=least frequent)

Context	Action				
	A	B	C	D	E
	Reteach Concept/ Skill	Let Student Progress	Assign Additional Practice With Same Materials	Provide In-Class Tutoring	Assign Additional Practice With Different Materials
Any Paid Aides					
no paid aides	*1.7(72)	Δ 4.8(67)	2.6(69)	Δ 2.5(69)	*2.5(69)
some paid aides	1.3(153)	5.1(148)	2.6(145)	2.1(153)	2.0(152)
Teaching Experience					
1 - 7 years	1.6(65)	5.0(65)	2.5(65)	2.3(65)	Δ 1.9(65)
8 - 15 years	1.3(92)	5.2(89)	2.7(89)	2.1(91)	2.1(92)
$\geq$ 16 years	1.4(87)	5.0(80)	2.7(83)	2.5(85)	2.4(84)
Class Organization					
team-taught classroom	1.3(81)	5.1(78)	2.8(78)	2.2(79)	1.9(69)
self-contained classroom	1.5(143)	5.0(138)	2.5(141)	2.3(143)	2.1(142)
Grade Levels Taught					
K - 2nd grade	*1.2(97)	5.1(92)	2.5(95)	2.2(97)	2.1(96)
3rd - 4th grade	1.5(50)	5.2(50)	2.7(50)	2.4(50)	2.1(49)
5th - 6th grade	1.5(57)	4.9(54)	2.7(55)	2.3(56)	2.3(56)
TOTAL SAMPLE	1.4(225)	5.1(215)	2.6(218)	2.3(222)	2.2(221)

*One-way analysis of variance significant at $\alpha \leq .05$.

Δ One-way analysis of variance significant at $.05 < \alpha \leq .10$.

Number responding are in parentheses.

Compared to teachers with more than 16 years of teaching experience, teachers with 1 to 7 years of teaching experience were significantly more likely to assign additional practice with different materials ($\alpha \leq .10$). They were also more likely to provide in-class tutoring and assign additional practice with the same materials (although the difference was not statistically significant).

Lower-grade teachers were significantly more likely than upper-grade teachers to reteach the same concept or skill. This result may be due to the higher incidence of aides in lower-grade classrooms.

The uses made by teachers of the information provided by required test information are presented in Table 11. Teachers most often used required test information for reporting to parents and/or staff and for evaluating the effectiveness of teaching methods and materials; consistent with other findings in the literature,

Table 11. How Teachers Use Information From Required Tests

Frequency of Use for Different Purposes
(1=never, 4=always)

Context	Purpose			
	A To Make Instructional Decisions	B To Evaluate Effectiveness of Methods/ Materials	C To Assign Grades	D To Provide Information To Parents/ Staff
Any Paid Aides				
no paid aides	1.9(67)	2.0(67)	1.3(67)	*2.3(68)
some paid aides	1.8(143)	1.9(143)	1.4(140)	2.0(139)
Teaching Experience				
1 - 7 years	1.7(60)	*1.8(60)	*1.2(60)	Δ 1.9(60)
8 - 15 years	1.9(88)	1.9(88)	1.4(87)	2.1(87)
$\geq$ 16 years	1.9(81)	2.1(81)	1.4(79)	2.2(78)
Class Organization				
team-taught classroom	1.9(69)	*2.2(69)	1.4(67)	*2.3(68)
self-contained classroom	1.8(139)	1.9(139)	1.4(139)	2.0(138)
Grade Levels Taught				
K - 2nd grade	1.8(86)	1.9(86)	1.4(86)	2.1(86)
3rd - 4th grade	1.9(48)	1.9(48)	1.4(47)	2.0(46)
5th - 6th grade	1.8(57)	2.1(57)	1.3(55)	2.2(56)
TOTAL SAMPLE	1.8(210)	2.0(210)	1.4(207)	2.1(207)

*One-way analysis of variance significant at $\alpha < .05$.

Δ One-way analysis of variance significant at $.05 < \alpha \leq .10$.

Number responding are in parentheses.

however, these frequencies were quite low. Required test results seemed to function as a standard of comparison as opposed to teacher-made tests which were used more for instructional decisions. Teachers without aides were more likely to use required test results to provide information to parents.

Findings with Regard to Teachers' Opinions and Knowledge. Teachers' opinions of factors influencing required test results are shown in Table 12. Teachers felt that the factors having the greatest influence on student scores on required tests were students' test-taking skills, student motivation, the quality of the test, and unusual circumstances (e.g., distractions, special activities scheduled). Parental influence was perceived as least influential.

Table 12. Teachers' Opinions of Factors That Influence Students' Scores on Required Tests

(1=little influence, 5=great influence)

Context	Factors Influencing Test Scores			
	A Students' Test-Taking Skills	B Quality of Classroom Instruction	C Student IQ	D Student Motivation
Any Paid Aides				
no paid aides	4.3(73)	3.7(72)	3.6(73)	4.3(72)
some paid aides	4.5(151)	3.7(148)	3.7(152)	4.2(152)
Teaching Experience				
1 - 7 years	4.4(65)	*3.5(64)	*3.5(65)	4.2(65)
8 - 15 years	4.5(92)	3.6(90)	3.8(92)	4.2(92)
≥ 16 years	4.4(88)	3.9(87)	3.9(89)	4.4(88)
Class Organization				
team-taught				
classroom	4.4(81)	3.8(78)	3.8(81)	4.3(80)
self-contained				
classroom	4.5(142)	3.7(142)	3.7(143)	4.3(143)
Grade Levels Taught				
K - 2nd grade	4.4(97)	3.7(93)	Δ3.8(97)	4.2(96)
3rd - 4th grade	4.3(50)	3.7(49)	4.0(50)	4.2(50)
5th - 6th grade	4.5(57)	3.7(58)	3.5(58)	4.4(58)
TOTAL SAMPLE	4.4(224)	3.7(220)	3.7(225)	4.3(224)

*One-way analysis of variance significant at $\alpha \leq .05$.

ΔOne-way analysis of variance significant at $.05 < \alpha \leq .10$.

Number responding are in parentheses.

Table 12. Teachers' Opinions of Factors Which Influence Students' Scores on Required Tests (Cont)

(1=little influence, 5=great influence)

Context	Factors Influencing Test Scores			
	E	F	G	H
	Parent Interest	Quality of Test	Unusual Circumstances	Students' True Ability
Any Paid Aides				
no paid aides	3.3(70)	4.4(72)	4.2(73)	3.8(73)
some paid aides	3.3(149)	4.4(149)	4.2(149)	4.0(151)
Teaching Experience				
1 - 7 years	3.3(65)	4.4(64)	4.2(65)	3.8(65)
8 - 15 years	3.1(88)	4.4(91)	4.1(91)	3.9(92)
≥ 16 years	3.4(87)	4.3(86)	4.2(87)	4.1(87)
Class Organization				
team-taught				
classroom	3.4(78)	4.4(79)	4.2(80)	*4.2(80)
self-contained				
classroom	3.2(140)	4.4(141)	4.2(142)	3.9(143)
Grade Levels Taught				
K - 2nd grade	3.2(95)	4.4(95)	4.1(96)	4.0(97)
3rd - 4th grade	3.2(49)	4.3(49)	4.2(49)	3.9(50)
5th - 6th grade	3.4(55)	4.4(56)	4.3(57)	3.8(57)
TOTAL SAMPLE	3.3(219)	4.4(221)	4.2(222)	3.9(224)

*One-way analysis of variance significant at $\alpha \leq .05$.

† One-way analysis of variance significant at $.05 < \alpha \leq .10$.

Number responding are in parentheses.

Responses differed significantly according to teacher experience. Less experienced teachers perceived quality of instruction and IQ as having less influence on test scores than did more experienced teachers. There was no difference in responses between teachers with or without paid aides. Team teachers felt that scores were more influenced by students' true ability than by teachers in self-contained classrooms.

Teacher reactions to the discontinuation of required testing are shown in Table 13. Teachers generally perceived that other teachers would react favorably to the discontinuation of testing, that administrators would have neutral feelings, and that parents' reactions would range from neutral to negative. Teachers' responses differed, depending on the presence or absence of an aide and on the teacher experience level. Teachers with aides reacted more positively to discontinuing testing than did teachers without aides. Less experienced teachers perceived administrators as having a more negative reaction toward discontinuation

Table 13. Teachers' Opinions on Reactions To Discontinuing Required Testing

(1=negatively, 5=positively)

	Parents' Reaction	Teachers' Reaction	Administrators' Reaction
Any Paid Aides			
no paid aides	2.4(72)	*3.5(72)	2.7(70)
some paid aides	2.5(153)	4.0(151)	3.0(145)
Teaching Experience			
1 - 7 years	2.5(64)	3.7(64)	Δ 2.5(62)
8 - 15 years	2.5(92)	4.0(91)	3.0(88)
$\geq$ 16 years	2.5(89)	3.7(88)	3.0(85)
Class Experience			
team-taught classroom	2.5(80)	3.6(80)	2.8(77)
self-contained classroom	2.6(144)	3.9(143)	2.9(137)
Grade Levels Taught			
K - 2nd grade	2.6(98)	*4.0(98)	3.0(94)
3rd - 4th grade	2.5(50)	3.8(49)	3.1(48)
5th - 6th grade	2.3(58)	3.5(58)	2.8(57)
TOTAL SAMPLE	2.5(225)	3.8(223)	2.9(215)

*One-way analysis of variance significant at $\alpha \leq .05$. Δ One-way analysis of variance significant at $.05 < \alpha \leq .10$.

Number responding are in parentheses.

of testing. Upper-grade teachers were more likely to rate parents' and administrators' reactions as negative than were lower-grade teachers. It should be noted that these latter teachers do more required testing.

Teachers were also asked to interpret percentiles and grade-equivalent scores, to give their opinion on the amount of time spent testing, and their opinion of the influence of classroom instruction on required test scores. Because of the similarity in the answers to these questions across the four groupings, only grand mean results are reported.

In response to questions about interpreting test results, only 50 percent of the teachers were able to interpret correctly percentiles and grade-equivalent scores (the two most common methods used for reporting test scores) or to describe the factors that influence test performance. Thus, although most teachers reported some training in measurement, half the questioned teachers were not able to interpret correctly standard test results.

Teachers were asked to rate the amount of classroom time spent in required test taking in relation to teacher/student benefits received on a five-point scale ranging from 1=too much to 5=not enough. Teachers were neutral; mean score for the 223 respondents was 2.3 from neutral.

When asked to rate the influence of instruction on required test scores on a scale of one (little effect) to five (great effect), mean response was moderate, 3.9 percent. Although teachers felt that tests to some extent reflected their instructional efforts, they felt that student test-taking skills and motivation and test quality were more influential.

Conclusions and Recommendations

This study sought to investigate two major areas of test use in schools: the frequency and types of tests administered in schools, and the major contextual variables affecting test use. A questionnaire was developed and administered to a stratified random sample of teachers in 20 California schools. Two hundred and sixty useable response sheets were returned. Teachers reported using many tests, both required and non-required, with program-embedded tests and teacher-developed tests being the most frequent forms of testing. Teachers also reported, however, that they were more likely to use interactions with students and personal observations than test results to assess students' progress.

Grade level was the major contextual variable that affected testing frequency. Although all grade levels were tested to some degree, the study confirmed the work by Goslin (1967) and Yeh (1978), which noted that there is less testing at the primary level. The survey also indicated that upper-grade teachers were more likely to use test results in their instructional planning.

The presence of classroom aides was another variable that appeared to affect test use. Teachers with aides were more likely to use test results. This finding may be a function of the instructional alternatives available to teachers with this kind of assistance. Consistent with this finding, teachers with aides were less likely to allow failing students to progress and more likely to provide students with tutoring or additional practice.

Teachers indicated that test-taking skills, test quality, and student motivation were more important factors in students' test scores than were quality of instruction or student ability. These perceptions may result from teachers' opinions that required tests do not reflect the content of their instructional program.

The study also indicated that teachers need more knowledge about measurement. Although a majority of teachers reported having some kind of training in testing, only about 50 percent could accurately interpret percentile scores or grade equivalent scores -- two rudimentary standardized testing concepts. It is therefore recommended that, if standardized tests are to be required, efforts be made to ensure proper interpretation of the results.

In summary, the study indicated that standardized testing comprises only a small fraction of the assessment activities. Curriculum-embedded and teacher-made tests are not only more prevalent, but also play a large role in student assessment. These are two areas that need additional educational research and training activities.

This study was sponsored in part with funds from the National Institute of Education under a special institutional agreement. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement should be inferred.

This paper is based on an earlier report prepared while the senior author was with the Center for the Study of Evaluation.

5. USES OF TESTS IN SCHOOL DISTRICTS

Mary M. Kennedy, The Huron Institute

In September 1979, my colleagues at The Huron Institute and I were awarded a contract by the National Institute of Education and the U.S. Office of Education to study the uses of evaluation and testing information in public school systems. The idea was to purposefully select school districts that were purportedly making good and appropriate use of such information; and to engage in rather lengthy, open-ended conversations with people in those school systems to learn not only what they were doing but also why -- why were they able to do what conventional wisdom says is not being done in the majority of school districts?

We planned a broad study, and this meant three things. First, we included a variety of kinds of information in our definition of "evaluation and testing" so that we could learn about the value of such data sources as survey results, enrollment statistics, and so forth. Second, we perceived our interviewees not as data users but as problem solvers, which meant that, in order to find out how people used data, we had to find out what kinds of problems they were trying to solve, for it would be in the process of solving these problems that they would use data. Finally, we imagined that the people who used data could be located almost anywhere in the system and could be trying to solve almost any kind of problem. For convenience, we grouped our potential users according to the types of issues they would be most likely to face: policy issues, program issues, or clinical issues (this last category refers to those issues addressed by teachers in their classroom activities with children).

We now have some 3,000 pages of transcribed notes. To simplify things a little, I will confine my remarks to the uses of test data. As it turns out, our convenient grouping of users is not completely appropriate to a discussion of test uses for two reasons. First, there are few if any examples of test data being used to develop any policies. And second, there are a number of ways in which test data are used that do not relate either to clinical or program issues, but could be called "systemic uses." So my three categories for discussing test use are these: independent teacher uses, systemic uses, and program directors' uses. They occur in districts that have adopted system-wide policies regarding the use of tests for instructional management and accountability. Although teachers are the primary people who implement systemic test-use programs, their use of tests under these circumstances is sufficiently different from the use of tests in districts without such programs that they justify a separate discussion.

Independent Teacher Uses of Tests

As I mentioned before, our strategy for finding out how people used data was to first find out what problems they were trying to solve. For teachers, there appeared to be three overriding issues: meeting children's social and emotional needs, meeting their instructional needs, and deciding when particular children need to be referred for special programs, such as gifted-and-talented programs or special education programs. These are not the only issues teachers discussed, but they were the most prominent. Others that came up occasionally, for which test data were used, included dealing with irate parents, motivating students, and coordinating lesson plans and schedules with other teachers.

As you might expect, very few teachers used test data when they addressed children's social or emotional needs. Most teachers who discussed this issue were not concerned about exceptional children who might need psychological assessment, but rather about ordinary children, each of whom had his or her own set of needs, motivations, and so on that teachers tried to take into account when they interacted with children. The primary issues for which test data are useful related to students' instructional needs and referral decisions. But even within these areas, teachers have available a variety of non-test data that they can draw on to make instructional decisions. Teachers give weekly quizzes, they have oral quizzes and recitations in class, they observe their students daily, and they take note of the kinds of questions students ask and the quality of their homework and recitations. The data they gather from these several sources contributes to their "working knowledge" of their students. It is not quantifiable, and it evolves as the school year goes on. It is more analogous to a "gestalt" of each child and of the classroom as a whole. Recognizing the role that working knowledge plays in teachers' instructional decisions helps us to understand the role that tests tend to play. There were three kinds of instructional decisions that seemed to come up repeatedly -- (1) grouping students, (2) on-going instructional decisions, and, of course, (3) referring students to other programs. Each topic involved the use of test data, but these uses are best understood in relationship to teachers' working knowledge.

(1) At the beginning of each school year, teachers make one of their most significant instructional decisions: how to group students for instruction. Yet, despite the significance of this decision, it is made before teachers have developed the body of knowledge that will guide them through most of their other instructional activities, and, consequently, they are more likely to rely on data from standardized tests. Although secondary teachers rarely group their students as explicitly as do elementary teachers, they nevertheless often try to identify those who may need extra help so that they can give the supplementary reading material or extra attention as the school year proceeds.

The type of test data teachers rely on at the beginning of the year depends a great deal on what is available, but assuming that a variety of test data were available, I think I could define certain preferences. First, when the course content is continuous across grade levels, as it often is in elementary reading and math, teachers rely most heavily on curriculum-embedded test data. For example, when an elementary school uses a single publisher's reading series for all grades, teachers want to know where students left off in the series the year before. They use the tests provided by the textbook publisher to determine which page number of which reader each child should begin the year on, and they group their students according to where different children are in these readers.

Not all schools use such series, however. Some have different textbooks at each grade level or even in each classroom within a grade level. In these cases, elementary teachers are in situations analogous to those of secondary teachers, whose courses are less often coordinated with other secondary courses. When coordination is not an issue, teachers turn to norm-referenced tests, not because they are interested in the norms per se, but because they perceive these tests as estimating students' general abilities and, hence, as useful guides for grouping or identifying students who may have problems keeping up. In fact, the emphasis on using these tests as ability tests is particularly evident in secondary education, where teachers may not even be interested in test performance in the content area they teach. For example, a social studies teacher may feel that the norm-referenced

test of social studies achievement has so little relationship with the curriculum that it is not useful, but that the reading test is useful. Since a great deal of the work in studying history or political science involves reading, knowledge of students' reading achievements may provide a reasonable prediction of their ability to do well in social studies.

(2) During the course of the school year, teachers rely primarily on their continually enduring working knowledge to make the on-going decisions they need to make. There are two conceptual problems with using standardized tests to determine students' instructional needs. The first is that the list of all possible needs is infinite, and no test can assess them all. To some extent, then, the instructional "needs" that are considered by teachers are those that tests happen to measure. The second problem is that teachers can only address needs they are capable of influencing, that is, those for which textbooks or other materials are available, or for which the teacher feels he or she has an adequate technique for teaching. Thus, a major influence on the use of test results is the extent to which the skills or knowledge a test happens to measure overlap with those that the teacher's materials and skills happen to cover. Of course, the overlap between these two sets of instructional tools is not entirely haphazard, but the more overlap occurs, the more useful the tests are. There are many teachers for whom only the tests they prepare themselves have the overlap they need.

When overlap is not high, as is often the case with norm-referenced tests, teachers tend to use the tests, as I mentioned earlier, as a general gauge of students' abilities, and during the majority of the school year, they don't need that guide -- they can rely on their working knowledge of their students. Indeed, teachers rarely seek out these data after the first 2 weeks or so of the school year.

But they do frequently receive these data during the course of the school year. Every district we visited had some form of a norm-referenced testing program and routinely sent test results back to teachers. Even though these data were unsolicited and tended to arrive late in the school year, teachers did look at them and used them to double check the professional judgments they had made of their students. And, when the two assessments -- that is, the test results and their working knowledge -- disagree, teachers are more likely to accept the higher assessment. Suppose, for example, a teacher has a student who seems to be doing reasonably well in class, but the student gets a low test score. In such a case, teachers are more likely to attribute the low test score to extraneous influences. If, on the other hand, the test score is higher than expected, they are more likely to reconsider their appraisal of the student, and often they give the student other tests as a means of double checking. These reassessments may not always lead to immediate changes in instruction however. If a child has been placed in a low reading group, for example, the teacher is not likely to suddenly move the child to a higher group. Instead, he or she may simply make a note to try to move the child faster through the workbooks.

(3) There are occasions during the course of the year when teachers may feel that their knowledge of students is not sufficient to make certain judgments. These occasions arise when a particular student's behavior is sufficiently different from that of other students; then the teacher may feel inadequate to interpret the child's behavior. Part of the reason teachers feel less comfortable with such students is because they are less familiar with them, but part of it is also that these deviations can lead to major decisions, such as referring a

child to special education, Title I, or a program for the gifted and talented. Since these children are unusual, teachers may seek test data to help them understand the children, but in most cases, they also seek other professional judgments.

Referral decisions are further complicated by the fact that each special program has its own set of eligibility rules. Title I services, for example, are offered only in those buildings with high concentrations of poor children. Within these buildings, eligibility is determined primarily on the basis of tested academic achievement. Special education, on the other hand, is supposed to be offered to any child who needs it, regardless of the building, but eligibility is much more difficult to determine. Not only must it be based on multiple assessments, but it must be determined by a multidisciplinary team that includes both the child's teacher and the child's parents. Many of our interviewees embraced the principles implied by the special education rules, arguing that decisions to place children in special programs were serious decisions and shouldn't be based on a single test or a single individual's point of view. They seemed less comfortable with referring children to Title I and to gifted and talented programs, for these programs did not have the elaborate procedures that special education had. Teachers felt tests were not sufficient to single-handedly make determinations about children's abilities or needs for these programs and appeared to feel that the rules did not allow professional judgments to be weighed along with the test data. It's not clear from our data that these requirements actually influence teachers' uses of tests. Right now I would say teachers continue to do what they think is in the best interests of the child, and if that conflicts with their perceptions of the regulations, they feel guilty but continue doing it.

Parents add yet a further complication to referral decisions. Parents are more likely to nominate their children for gifted and talented programs, rather than waiting for teachers to do so, and are more likely to reject the notion that their children need special education. Teachers have little training in how to discuss children's capabilities with parents, and the subject is a very sensitive one for parents. Parental pressure can cause teachers to rely more heavily on test data in two ways. First, teachers often feel they can justify their position better with test data than with their privately developing working knowledge. The data provide a "shorthand" method for communicating with parents and they are perceived by parents as being more objective and more authoritative than a teacher's professional judgment. Second, if parents raise counterevidence, based on their judgment as parents, they may cause teachers to lose confidence in their own judgments -- confidence that is already shaken by the fact that the child is unusual -- and thus cause teachers to rely more heavily on test data as well as on the professional judgments of others to understand the child.

To summarize what we have found with regard to teachers' uses of tests, then, the predominant uses occur when teachers have relatively little professional knowledge with which to make decisions. The two primary times when this occurs are at the beginning of the year when teachers don't know their new classes, and during the year when a student is sufficiently different from the norm that his or her behavior is hard to interpret. For the most part, standardized tests appear to be considered as secondary sources of information, rather than primary. Teachers' primary source -- their working knowledge -- gives them information about children's performance over a longer period of time and over a much wider range of situations than do test data.

Systemic Uses of Test Data

So far, my remarks have been limited to what teachers do within the confinement of their own classrooms, when there are no district-wide pressures to use test data in any particular way. You probably also have noticed that I have not discussed criterion-referenced tests. As it turns out, these latter two topics (district-wide policies and criterion-referenced tests) go together. Although it has now become conventional wisdom that one can't easily distinguish between norm-referenced and criterion-referenced tests -- and our findings confirm that wisdom -- one can easily distinguish between the uses of tests for district-wide instructional management and uses of tests by individual teachers. Many of our school districts have established district-wide testing systems and associated rules for their use, which for convenience I will label as criterion-referenced systems.* The main point that needs to be made about these systems is that teachers' responses to them, particularly their responses in terms of how they use the test data, are more a function of the rules associated with the tests than they are a function of the design or nature of the tests themselves. Rather than taking the time to describe several examples of such systems, the incentives they introduce, and teachers' responses to them, I will try to give you instead three general principles about these systems that my staff and I are fairly sure hold up.

1. What teachers seem to like most about these systems is that they determine for teachers what their instructional goals should be. The systems focus teachers' attention on specific skills or knowledge that students should have and often organize these skills into clusters. This may seem like a trivial use, but it is not. The goals of education are extremely broad and vague, not very amenable to organized attack. Furthermore, teachers are faced with classrooms of real students, who vary not only in their abilities but also in their interests and motivations. The task of organizing a diverse group of youngsters to move toward a diffuse goal can be overwhelming and can leave teachers disorganized and undirected. They can spend more time responding to students' ad hoc behavior than in initiating activities of their own. Given these circumstances of classroom life, a test that reports children's progress on specific goals at regular intervals can be a major force in shaping teachers' behavior. It reminds them at regular intervals what they should be striving toward and how much progress they've made so far.

2. The counterside of the first point is that these tests can focus teachers' attention on superficial goals, or can encourage them to raise test scores without regard for the underlying cognitive processes that should be affected. Test scores are, after all, only indicators of student characteristics, and unless teachers are regularly reminded of that, they can easily slip into an assumption that the scores are all that matter. This is particularly true when building principals or other supervisors in the system place a heavy emphasis on progress in test scores and assess teachers' capabilities on the basis of their children's test scores. We visited one school district in which students learned nothing but grammar, from elementary language courses all the way through secondary English classes, and another in which elementary reading teachers focused on phonetics even with students who could read fluently. In both cases, course content was driven exclusively by the district-wide testing program. Thus, even though these systems can greatly benefit teachers by focusing them toward a more definable educational goal,

*Not all systems, however, rely on criterion-referenced tests. Some use norm-referenced tests but establish performance criteria for them.

they can also hinder teachers if the test content is perceived as dictatorial. Unfortunately, the line between these two ways in which system-wide testing policies affect teacher usage is extremely delicate, and a school system can skip over it almost unconsciously.

3. My third point about these systems is that they are rarely really individualized systems, primarily because teachers do not have the resources to teach children according to their individual needs. Teachers may use the data to re-group children according to their current skills and deficits, but instruction continues to occur in groups. In those districts or classrooms where individualized instruction occurred, teachers had ample assistance, either from aides or parent volunteers, and claimed that the recordkeeping involved demanded extensive staff time. In one district, for example, teachers claimed they needed 2 hours each day to prepare for a 20-minute reading class!

System-wide testing policies, then, do change teachers' uses of tests. Without them, teachers rely primarily on their own professional judgment of students and try as best they can to organize their diverse groups of students to move toward broad and vague educational goals. They rely on norm-referenced tests primarily at the beginning of the year, or when they suspect that a child has special needs. Under system-wide policies, teachers are more likely to engage in systematic instruction and to focus that instruction on the skills or knowledge measured by the test. For the most part, we had the sense that they didn't need the test scores per se -- they could have estimated children's abilities using their professional judgment -- but did use the test content to determine what their instructional goals would be, and in some cases felt that their professional status or even their job security rested on their children's performance on these tests.

Program Directors' Uses of Tests

Those individuals whom we grouped together into the category of program directors actually held a variety of different positions. Some managed categorical programs such as Title I or special education, some managed district-wide curricula, such as elementary reading or secondary social studies, and still others managed special district programs such as preschool programs or basic skills programs. These are the individuals toward whom most formative and summative evaluations are aimed. Before commenting on their uses of test data, I should point out the major concerns that they had. There were four, and they all influenced program directors' use of test data. First, most program directors felt they did not have the authority they needed to meet their responsibilities. Though they may have been responsible for assuring compliance with program regulations, or for assuring that a curriculum was implemented, they did not supervise building principals and rarely did they directly supervise program teachers. Second, for those who managed federally or State-funded programs, compliance with associated rules and regulations was a major concern. Third, whether their programs were supported by external or local funds, program directors were concerned about their funding level. They were rarely in a position to influence funding very much, since they competed either with other local programs or with like categorical programs in other districts. They were often either unaware of the competition or were not privy to the decisionmaking process that would ultimately allocate funds across competing programs. Finally, some program directors discussed program improvement as an issue, though this concern seemed less frequent and was almost always entangled with the other three issues.

The first three of these concerns leave most program directors feeling either too distracted by funding and compliance concerns to worry about program quality or too powerless to feel they can affect program quality. Nevertheless some did concern themselves with ways of improving their programs. And those who did often used test scores to address that concern.

Test scores are, of course, a major part of summative program evaluations, but these are generally done only for externally funded programs and are rarely if ever attended to by program directors. Instead of attending to these studies, program directors tend to engage in a variety of ad hoc analyses of patterns in test scores. A director might, for example, compare scores in reading with those in math, or compare social studies scores across buildings or grade levels. Sometimes they follow up these comparisons by doing some on-site investigations of their own to see if they can figure out why differences exist. Keeping in mind, however, that their authority to make changes is limited, their budget is set elsewhere, and that any changes they make have to be consistent with regulations and union agreements, they have very few options for acting upon any findings they may feel need to be acted upon. In fact, the main strategy they have for improving programs is staff training.

The uses that program directors make of test data, then, differ from the uses teachers make in several respects. First, program directors are more interested in patterns of scores, as a means of determining the content and audiences for possible staff training sessions, than they are in individual scores. Second, since program directors do not have the day-to-day observations of students that teachers have, they are more likely to accept test scores as their primary source of evidence of how much children have learned. Third, rather than using scores to reassign or regroup children, they are more likely to use them to make decisions about teachers and their instructional needs. Finally, I should point out that program directors also have a large store of their own working knowledge; they observe classrooms, meet regularly with principals, instructional coordinators, resource teachers, and parent groups, all of whom have also observed classrooms. In addition, they meet with district administrators above them in the hierarchy, and often meet with State and Federal staff members as well. The breadth of these contacts means that they have an evolving gestalt of their program, and they use that knowledge to determine how test data will be used. That knowledge allows them to estimate such things as whether a desired program change can be afforded or would comply with existing rules, whether a desired staff training program would be agreed to by the teachers union, whether a desired reassignment plan would be agreed to by the district administrators, and so on.

The reason I emphasize the role of the program directors' working knowledge in conjunction with test use is that a lot of people seem to feel that program directors use data for primarily political purposes; that is, to advocate their program, demonstrate the need for funds, and so on. I am not so sure that these uses are "merely" political. When one considers the constraints within which program directors must operate -- their lack of authority, their lack of influence over their funding, and the several regulations that they must often comply with -- these political uses are necessary if program improvements are to occur. In most "use" situations, there is both a program improvement motive and a power struggle motive involved. The two are hard to separate. Hence, although program directors may place more faith in the validity and accuracy of test data than teachers do, they still rely heavily on their working knowledge to translate those data into action.

Summary

I have tried to list some of the ways in which we found test data being used in the 18 school districts we visited during 1980. The uses seem to be fairly different depending on whether we are talking about a teacher in isolation, a teacher in a district with a system-wide testing policy, or a program director. I hesitate to claim that these are definitive statements, since our data have not been thoroughly analyzed. Instead, they are today's best guesses about what is going on. Tomorrow's may be slightly different.

6. INTEGRATING ASSESSMENT WITH INSTRUCTION¹

Herbert C. Rudman, Michigan State University

I should like to thank the National Institute of Education (NIE) for inviting me to address this conference and specifically to thank Jeffery Schiller, Judy Shoemaker, and Lawrence Rudner for their assistance and support to our team through this first year of our work. Our appreciation is also extended to the Department of Defense Dependents Schools (DODDS), who have cooperated with and cofunded this project. I especially wish to take this opportunity to thank Dennis Holmes, Steven Motta, and Janet Wall for their suggestions to our staff through much of our work.

I speak this morning as the Director of the Assessment project, but its deliberations, decisions, and activities are the result of four outstanding colleagues: William A. Mehrens, Donna S. Wanous, Christopher Clark and Andrew Porter. You can't see them, but they are here in every word I speak.

Let me share a road map of the next hour with you. The project reported here is a response to an RFP that was issued jointly by NIE and DODDS. Our activities in 1980 have sought to address the gap that exists between testing and assessment on the one hand and classroom instruction on the other. Our team addressed two goals this first year; we have attempted to identify the critical concepts needed in the interpretation of classroom-related data, and also those critical "decision-points" that occur in teaching, and we have identified three vehicles for imparting these concepts to teachers.

It is my intention this morning to share with you what we have learned to date about teacher use of tests and other information-gathering devices, and what we have learned about the link that presently exists between testing and teaching. We will examine some principles we have drawn concerning in-service programs for teachers, and we will briefly report on some of the help presently available to teachers through the public and private sectors.

I shall also share with you the highlights of a review of the literature undertaken by our team; a review that covered a 60-year sweep of the research literature available to us through conventional and less-than-conventional sources. We went into this sweep confident that the collective experiences of this team would mirror what the literature would reflect; but there were a few insights gained that we had not anticipated.

And, finally, I shall share with you what we have developed to date, the components of the training system and how we propose to package and use them with teachers.

A comprehensive sweep of the literature reveals that the integration of assessment and instruction is a "land" that has been visited and revisited many times by many educationists, researchers, sociologists, critics of schools, and educational apologists. The result is a literature filled with inconsistencies

¹Based on Herbert C. Rudman, Janet L. Kelly, Donna S. Wanous, William A. Mehrens, Christopher M. Clark and Andrew C. Porter, 1980.

that is uneven in quality and in scope of serious study. The unwary can easily fall prey to the belief that what they write is new, unique, and definitive.

To complicate matters still further, much has been done over the years -- and is being done -- that has not yet found its way into the traditional channels of literature (reports of scientific meetings, articles published in research journals and professional magazines, books, and the like). A major bibliographic tool is Educational Resources Information Center (ERIC), and yet ERIC is severely limited in its review and reporting of copyrighted material available to help teachers use assessment materials. Much of ERIC's recent literature (1963 to the present) deals with the "fugitive" materials, papers given at scientific meetings, internal professional reports, unpublished studies, and the like. A major effort has been made to give this type of information wider circulation through ERIC. But if one is to get a full picture of the linkage of assessment with instruction, one needs to examine what is and has been available to teachers and other educational practitioners that would make such linkage more feasible. My colleagues and I have attempted to capture some of the commercially available material that teachers can use in interpreting test results and in translating these results into classroom practice.

The Setting

Teachers have, for many years, been testing students with a variety of instruments, some standardized, and many teacher-made. But testing for the most part has been summative. It has been used to rate pupil progress rather than to monitor it. There is little evidence to indicate that teachers use their own tailor-made tests analytically to identify modalities upon which to build learning experiences. Testing, while not new to the educational process, has been separate from it.

There is a growing sense of need today to link assessment with teaching and to make instructional decisions more rational and perhaps less intuitive. How much do we really know about the linkage that presently exists between assessment and instruction? How do teachers make instructional decisions?

There is no good substitute for teacher intuition and judgment. Without intuition, teaching is lean and somewhat mechanical, but that intuition can be enhanced and teaching enriched when information is added. A problem facing all of us in education is that we tend to separate testing from teaching. Unless teaching is closely linked to testing, one could conceivably "over-test" with the administration of only one formal examination if the information gleaned is never used by the teacher.

In examining the literature that specifically addresses teacher use of data and more general questions of how teachers make classroom-related decisions, Clark & Yinger (1977) report a paucity of information and a new interest in the topic.

It seems that those who exhibit considerable skepticism of assessment techniques in general, and standardized tests in particular, tend to view testing and teaching as two separate events, which have little or no relationship one to the other (Bhaerman, 1977; Wise, 1978; Shulman, 1978). Whether assessment and testing have no relationship to instruction because of the inherently poor construction of some standardized achievement tests (McKenna, 1977; Rudman, 1970, 1971, 1973, 1974, 1975, 1977), or because of inappropriate use of these devices (Rudman, 1976), may be a moot point. For whatever the reason, there appears to be a widespread belief

that teachers are skeptical about the values of external assessment information yielded by tests not devised by the teachers themselves (Perrone, 1977).

A persistent notion that testing takes time away from teaching (Shanker, 1979) is the result, most likely, of an inability to identify the symbiotic relationship that can exist between instruction and assessment. There appears to be no corresponding skepticism about the relationship between planning and instruction. Clark and Yinger, in surveying the literature on teacher planning, report that considerable attention has been paid to the need for planning in instruction. However, they report that "empirical studies of planning have only been conducted since 1970, and to date the published studies can still be counted on one hand." (Clark & Yinger, 1977, p. 281).

A preponderance of teacher contracts calls for planning time to be allocated during the school day. Given the implicit recognition of planning as a legitimate function of teaching, it would appear that too little attention has been placed in teacher education curriculums on the need for developing a similar understanding of the relationship between data gathering, planning, and instruction. (Hastings, Runkel & Damrin, 1961; Farr and Griffin, 1973; Durost, 1959; Roeder, 1972).

A group that has been active in highlighting the issues connected with evaluation of classroom instruction has been the National Consortium on Testing. The Consortium is composed of representatives of various special-interest groups. One of these groups, the National Educational Association, has called for a moratorium on testing (1974). Some of the Consortium's membership have taken a more moderate stand (e.g., The National Council of Teachers of English and the American Federation of Teachers, 1976), and others have simply said that assessment in general, and standardized testing in particular, is a vital and necessary part of instruction (e.g., representatives of the National Council on Measurement in Education and the American Education Research Association).

Considerable discussion has appeared in the literature concerning the impact that standardized testing has had upon the attitudes of pupils, parents, teachers, school board members, and the general population toward school learning. A survey conducted among 1,022 school board members and their administrators (NSBA, 1977), revealed that school board members generally believe the following:

1. Standardized testing programs are worth the required investment of time and money.
2. Their teachers and administrators understand test results.
3. It would not be wise to eliminate standardized testing.
4. Those who call for the elimination of assessments are only trying to avoid being held accountable.

In one of a series of studies conducted by Stetz (1977) and Stetz and Beck (1978, 1979), teachers and administrators were asked to respond to a semantic differential concerned with the mood of parents and teachers toward standardized achievement testing. Teachers tended to view themselves as knowledgeable about testing, interested in the results yielded by tests, somewhat anxious about the results, and generally supportive of testing.

Much has been written of the "Pygmalian Effect" upon teachers' perceptions of their pupils' performance and the resulting effect upon actual student behavior (cf. Rosenthal and Jacobson, 1968). This issue was reexamined by others who conducted a longitudinal study of tests used in Ireland (Pedulla, Airasian, Madaus and Kellaghan, 1977). Several of the findings of that study have significance in the linkage of assessment and instruction:

1. Teachers form strong and stable expectations for their students early in the school year independently of any standardized test results.
2. Changes that do occur in teacher expectations have a "halo" effect. If teachers raise their expectations of a pupil's performance in English, they tend to carry over this estimate into mathematics as well.
3. Teachers' expectations are more global than subject-specific. Students are viewed as "smart" or "slow" generally, rather than able in one area and less so in another.
4. In very few instances did test results alter teacher expectations of their pupils (10 percent or less). In most instances when changes in estimates did occur they tended to increase rather than lower teachers' perceptions of pupil achievement.

In general, then, this study seemed to conclude that there is little or no "Pygmalian Effect" through the use of standardized test results.

The period 1974 to 1977 saw considerable acrimony exhibited toward the use of tests in the classroom (Rudman, 1977). Since 1977 more constructive discussion has emerged concerning the uses, the advantages, and the disadvantages of standardized tests and various other assessment techniques and devices in classroom instruction. (Freeman, Kuhs, Knappen, and Porter, in press; Shulman, 1978; Mehrens and Ebel, 1979). In August 1978, the National Institute of Education sponsored a conference on research in testing, teaching, and learning. Those who presented papers at this conference represented a diverse group. At the end of the meeting, three sets of recommendations emerged: the need for a better fit of testing to the cultural backgrounds of students, the need for a better fit of testing to educational objectives pursued, and a need to combine testing with teaching (Tyler & White, 1979).

In March 1978, then HEW Secretary Joseph A. Califano, Jr., convened the National Conference on Achievement Testing and Basic Skills. Among the many conclusions and recommendations that emerged from that meeting were the following, which have particular relevance to the topic of this review:

1. Testing is part of a large consumer and accountability movement. One cannot talk about testing without talking also about the need for education to be accountable for achieving its goals.
2. Testing and evaluation are useful tools. They should not be eliminated merely because they are the messengers of bad tidings.
3. Education goals should drive the agenda of testing and evaluation and not the other way around.

4. Instead of asking how we can make better tests, we should ask how we can make better use of tests, how tests can be effective means of diagnosis, and how that diagnosis can in turn be translated into educational intervention, remediation, and compensatory efforts.
5. Research questions to be addressed included the following:
 - a. What are the appropriate uses for each of the different kinds of tests?
 - b. How can the reporting of test results be improved?
 - c. How can tests be used to provide better diagnostic assistance for individuals and groups of individuals?
 - d. What do we know about the validity and effects of teacher-constructed tests?
 - e. What are the alternatives to testing for evaluation, diagnosis, and prediction? (Tomlinson & Treacy, 1979).

Decisionmaking is an everyday occurrence for classroom teachers and building administrators (Rudman, 1976, 1977). The decisions to be made range from the simple and mundane to those that are complex and vital. Some are made instinctively; others require considerable thought. The quality of goodness of the decisions depends in some measure on the process used to arrive at them (Clark & Yinger, 1977). Was the decision made because one "felt right" about it? Was a solution formulated because one recognized familiar elements in it and called on a former solution to meet the new problem? Not every problem demands extensive analysis and reflective thought. This is true of the simple problems. What of those, though, that are not so simple?

Analysis of complex problems calls for systematic study of the existing situation (Rudman, 1977; Hickey and Hoffman, 1973; Shavelson & Borko, 1979). It calls for the collection and interpretation of data. Information will not solve a problem, but it will serve as a basis for considering alternative solutions and allow a consideration of the consequences of these alternatives (Ebel, 1975).

Rarely do teachers find themselves confronted by simple problems in instruction. Instead they are faced with such complex variables as personality, social background, instructional competence, physical or mental health, motivation, and family, child, and school values. While the empirical studies reported to date are not encouraging in terms of the use of information made by teachers in the planning of classroom activities (Clark & Yinger, 1977), there is agreement among those who have reflected on the planning process that information is useful in an evaluation of the learning activity (Krug, 1950; Tyler, 1950; Taba, 1962; and Popham & Baker, 1970).

The fact that the studies reported to date reflect little use made of information by those teachers studied may well be an artifact of neglect on the part of teacher educators and building administrators. Some significant and important decisions are made by the teacher early in the year (grouping, pacing, remediation, time allocation, subject area weakness, and strength identification within a class), and all of these decisions can be enhanced by the proper utilization of information. There is some evidence to indicate that these decisions do indeed

come early in the year, within the first 3 or 4 weeks of the school year (Tikunoff and Ward, 1978; Buckley and Cooper, 1978; and Schultz and Florio, 1979).

If evaluation information is to be of much help to teachers, it must be available soon after the school year begins. "Timing is important, for these decisions will not wait, and they are decisions that have powerful shaping effects for the entire school year" (Clark, 1979).

Several researchers (Buckley and Cooper, 1978; Dasho, 1978; Schultz and Florio, 1979; and Tikunoff and Ward, 1978) have identified the first 2 weeks of school as a critical period in the classroom. During this time period teachers are making a variety of assessments about their students' abilities across several dimensions. It has been suggested that this is a critical point at which teachers could be reflective about their use of measurement concepts in their assessments, how these concepts relate to their instructional decisions, and about which concepts they would like further instruction.

Buckley and Cooper (1978) also found that the entrance of a new student into the classroom was a critical event for both the teacher and the student. Their research suggests that this is another time when teachers make important assessments of a student's abilities for future instructional placement.

The Focal Points of This Review

The remainder of this review will be organized around the types of decisions made by teachers in the classroom and around the concepts and skills needed by them to use and interpret information. Specifically, this review will focus upon four questions:

How Do Teachers Use Test Information for Instructional Decisions?

How often and in what way do they refer to norm-referenced, standardized achievement tests? Do they tend to use objective-referenced and criterion-referenced tests? Do they tend to use objective- and criterion-referenced standardized achievement tests more often than they do norm-referenced? To what extent are teachers knowledgeable about the interpretation of test information? In what ways do they use other assessment measures in making instructional decisions? What kinds of instructional decisions are made that use data as an important basis? Is there a difference in the decisions that are made intuitively as compared with those made using information? Are there observable differences in the consequences of decisions made using data as compared with those that are intuitively based?

What Has Research Indicated About the Linkage Between Testing and Teaching?

To what extent have we discovered a positive link between successful diagnoses of educational problems and effective treatment based upon diagnosis? Does research indicate specific and documented instances where the use of standardized diagnostic instruments have resulted in effective treatment of educational inadequacies? Which diagnostic tests have proven effective in diagnosis? Can standardized achievement tests (survey type) be used effectively for systematic analysis of learning? Are there some areas in the curriculum more amenable to this type of analysis than others? What does research

indicate about the successful use of tailor-made tests (written by teachers or by commercial contractors) to improve the quality of achievement of students?

What is Known About the Successful Practice of Inservice Training of Teachers?

Is there a consistent pattern emerging in those teacher education programs that appear to be successful? Can we identify those elements that consistently appear in programs identified as successful? What have we learned from the inservice preparation of practitioners in other fields, e.g., law, medicine, management? What work has been done in identifying critical incidents in teaching practice? Can these critical incidents be successfully simulated?

What Is and Has Been Available to Teachers to Aid Them in Interpreting and Using Assessment Information for Instructional Decisionmaking?

A considerable collection of materials is emerging that deals with the specifics of how to use assessment information in the building and in the classroom. State agencies that are involved in mandated assessment programs have developed manuals for teachers on the use of these data. Commercial publishers have, to varying degrees, developed manuals for teachers and administrators -- manuals designed to help them interpret and use the results derived from standardized tests. Some of these publishers have also developed audiovisual aids for inservice programs. Some local districts have developed their own data-collecting devices that cover a wide range of learning-related activities and have produced aids to teachers. Professional associations have also developed materials for teachers.

How Do Teachers Use Test Information for Instructional Decisions? That there is little new under the sun is brought home to the contemporary reader of an article written almost 60 years ago (Washburne, 1922). The author reported on work begun in the Winnetka, Illinois public schools in 1919. Carlton Washburne, Superintendent of Schools and former director of testing in the San Francisco State Normal School, described his approach to making educational measurement a key to individual instruction and academic promotions.

After reviewing the effect on education of the time/management movement of the early 20th century, Washburne said:

...As a consequence of making time the constant factor, rapid pupils either acquire habits of comparative idleness and work below their ability, or are pushed ahead a grade, losing much valuable intermediate instruction; slow pupils are either advanced with inadequate foundation or are retained abnormally by grade repetition....

Washburne then went on to describe the process of using achievement tests to help determine the child's progress through the grades; a process that came to be known as the Winnetka Plan.

Before either tests or practice materials could be prepared it was necessary to establish our units of achievement. We call these units, "goals." Wherever possible we established them upon the basis of research already done....

...each "goal" must be achieved by each pupil before he can go on to the next goal.

...Instead of reciting the children practice and prepare themselves for achievement tests. The teacher instead of listening to recitations passes about among the children, helping them in the preparation of their materials.

Sound Familiar? -- 1922.

How teachers use tests -- a summary. I posed a number of questions that we expected would be addressed somewhere in the literature that covered a 60-year span. Many of these questions have been examined over the years, although some of the materials reviewed are prescriptive essays rather than descriptions of classroom practice or experimental, laboratory-induced situations. In some instances the answers to these questions are consistent and even repetitive. In other instances the evidence is mixed. A disturbing element recently injected into the literature on educational measurement is that measurement specialists may not be accurate perceivers of the testing practice of non-measurement practitioners. Descriptions of teacher antipathy towards the use of tests have almost become the "coin of the realm;" yet data uncovered in this review would cast some doubt upon that generalization (Stetz and Beck, 1977, 1978, 1979; Feldhusen, 1964; Yeh, 1973; Olejnik, 1979; O'Regan, et al., 1979; Russo, 1978; Beggs, et al., 1972).

What have we learned about how teachers use tests? The following summary is a partial answer to that question:

How knowledgeable are teachers about test interpretation?

1. While teachers rather consistently report that they are comfortable with the scores they receive after the use of standardized tests (Stetz and Beck, 1977, 1978, 1979; Olejnik, 1979), when given interpretation items on a test, they misinterpret the measurement concepts presented (Olejnik, 1979; Nearine, 1970; Yeh, 1978; Goslin, 1965, 1967; Fleming, 1971).
2. Teachers' estimates of pupil performance on tests are remarkably stable and consistent (Morine-Dershimer, 1979; Sorotzkin, Fleming and Anttonen, 1974; Beggs, et al., 1972; Pedulla, et al., 1977); they are not nearly as prone to expectation set as earlier characterized (Rosenthal and Jacobson, 1968).

What are the attitudes towards testing of students, parents, school board members, teachers, and administrators?

3. It would appear that there is a gap between what those most closely affected by testing think, and what others write about what they think. (Stetz and Beck, 1977, 1978, 1979; Goslin, 1967; Rudman, 1977).
4. Students report that frequent testing helps them retain more content, increase test anxiety in a beneficial way, and helps them monitor their progress through a course. (Feldhusen, 1964).

5. Students reported a higher level of test anxiety over teacher-made tests (64 percent) than they did over standardized achievement tests (30 percent). While 75 percent of the students thought that standardized achievement tests were fair, only 38 percent of them wanted to take another one the following year. While 88 percent of the students wanted to know the results of their tests, only 44 percent wanted to discuss them with their teachers (Stetz and Beck, 1979).
6. Teachers generally seem to be much more supportive of testing than typified in some of the literature. They certainly do not reflect the antipathy toward testing that the National Education Association shows. (McKenna, 1977) Teachers' attitudes are in part a function of their experience and their knowledge of testing. Those who are more experienced and who have had some measurement training and experience in administering tests show greater support for the use of them than do those who are less experienced and less knowledgeable (Yeh, 1978).
7. School board members are at odds with the professional leadership of their national association. They tend to place much more credibility in the use of tests, and on the ability of their professional staffs to correctly interpret the results of tests, than the published materials of their associations would indicate (Rudman, 1977; NSBA, 1977; Bhaerman, 1977).
8. School superintendents report a higher level of satisfaction with the use of tests for decisionmaking than do measurement specialists, whose perceptions of others tended to be pessimistic (see Beck).

Are there differences in the frequency of use of norm-referenced, objective and criterion-referenced tests? Which do teachers seem to prefer?

9. Recent writing would indicate that the use of criterion-referenced tests is preferred over that of norm-referenced tests, but with the exception of one article reviewed, most of the literature calling for the use of criterion-referenced tests is prescriptive rather than descriptive (Yeh, 1978; Popham and Husek, 1967; Mills and Hambleton, in press; NEA, 1975; Airasian and Maduas, 1972).
10. Those studies that were descriptive tended to show a preference for norm-referenced tests and the standard scores in which the results are couched (Yeh, 1978; Olejnik, 1979; O'Regan, et. al., 1979; Nearine, 1970).
11. Only 35 percent of teachers quizzed favored increased use of criterion-referenced tests (see Beck).

What kinds of instructional decisions are aided by data?

12. Even though the primary thrust of our review has been on the linkage between classroom instruction and assessment, an interesting monograph surfaced that has some relationship to the broader topic. In a study conducted by the Bureau of School Services, University of Michigan, 13 uses of test data were identified (Brzezinski, 1979). The top five uses of test data by districts were identified as (1) reporting to boards of

education throughout the year on the progress being made in the interpretation of test scores (63 percent of the districts sampled), (2) relating other data to assessment information (51 percent), (3) conducting a study to determine which objectives were being taught in each grade (47 percent), (4) appointing curriculum committees to study test data and to relate them to current teaching strategies (40 percent) and (5) developing study committees to discuss the linkage between test results and curriculum in grades.

13. A planning model presented by Shavelson and Borko calls for estimates of student attitudes as an essential element in the planning of instructional decisions. "Teachers report that they use a wide variety of information in planning instruction for students." (Shavelson and Borko, 1979; Borko, 1978; Russo, 1978). The authors dwell especially on instruction related to the teaching of reading.

14. There is emerging in the research on how teachers think, a body of literature that indicates that teachers do not plan or make instructional decisions in quite the logical, rational pattern that those who teach them would prefer.

"In particular, the teachers studied did not begin or guide their planning in relation to clearly specified objectives or goals. Rather, teacher planning seems to begin with the content to be taught and considerations about the setting in which teaching will take place. The focus then shifts to student involvement as a process objective. The activity, rather than the objective seems to be the unit of planning." (Clark and Yinger, 1977).

15. Other studies reviewed here identify pacing, grouping, and placement in advanced programs as types of activities aided by test data (Stanley, 1978; Feldhusen, 1964; Marso, 1970; Jones, 1974; Washburne, 1922).

Finally, two questions initially posed at the beginning of this review must remain unanswered. "Is there a difference in the decisions that are made intuitively as compared with those made using information?" "Are there observable differences in the consequences of decisions made using data as compared with those that are intuitively based?" Nothing in the review of any substantive nature could be found that would address this question in relation to the classroom environment, whether experimentally or pragmatically based.

What Has Research Indicated About the Linkage Between Testing and Teaching?

Those who have studied the relationship between testing and teaching have addressed several areas, the linkage between diagnostic assessments and teaching interventions, the relationships between diagnosis and successful treatment of specific educational inadequacies, and the use and construction of tailor-made tests. But educational researchers are not the only ones interested in the construction, interpretation, and use of tests; legal scholars have also turned their attention to these topics (Mills and Bryan, 1977; Tractenberg and Jacoby, 1974; Nolte, 1975; Weckstein, 1973; Lerner, 1978).

Linking testing and teaching -- a summary. Much has been written and discussed concerning the logical use of information in the teaching process. Not all that has been written, indeed, only a small portion of it, has been research-based.

While conceptual schemes are most certainly the root of research, conceptual schema without a research base are limited vehicles for educational practice.

One of the most widely used models for rational, information-based teaching is the diagnostic-prescriptive approach. Its advocates are many and represent a variety of educational disciplines (Clark and Yinger, 1977; Rudman, 1977; Shavelson and Borko, 1979; Pedulla, et al., 1977; Beggs, et al., 1971; Bloom, et al., 1971). While the components of the diagnostic-prescriptive teaching model differ somewhat from advocate to advocate, they can be characterized as the gathering of information, the classification of students on the basis of the information, and an analysis of strengths and weaknesses to be addressed during the ensuing instruction. This process, in turn, calls up such psychometric concepts as validity, reliability, questions of bias, test construction issues, and the like. These concepts are the very ones that have been addressed in recent years by those who have come before the courts with grievances about the application of test results to themselves or to others for whom they appear as advocates (Mills and Bryan, 1977; Tractenberg and Jacoby, 1974; Nolte, 1975; Weckstein, 1973; Lerner, 1978).

What has the research literature to say about the law's interpretations of test use? What do we know about the research on linking assessment to teaching? The following summary is a partial answer to these questions:

What have recent court decisions concluded about the uses of tests?

1. To establish a challenge to test use one need only show that the testing procedures produce significant disparities between the percentage of the group represented in a particular classification and its percentage in the total testing population. There need be no evidence of discriminatory intent (Weckstein, 1973).
2. All those rejected or classified on the basis of test results regardless of race, class, ethnicity, or sex can attack classification procedures (Armstead v. Starkville Municipal Separate School District, 1972, as cited in Weckstein, 1973).
3. Testing is severely constricted in school districts that have recently undergone desegregation (Nolte, 1975).
4. Past discrimination in school districts has been judged to be evidence of the discriminatory use of tests (Weckstein, 1973).
5. Recent rulings seem to indicate that even when tests have been shown to be valid, their use is judged to be discriminatory if the characteristics being measured have been subjected to past discriminatory policies (Weckstein, 1973).
6. Tests can be used if they can demonstrate that by classifying students they can improve their previous status (Weckstein, 1973).

How sensitive are teachers toward the information they use in making instructional decisions? How accurately can teachers and parents assess student behavior? Are teachers good reporters of their own instructional behaviors?

7. There is conflicting evidence concerning teachers' sensitivity to the information they use in instruction. Some researchers (Shavelson, Cadwell, and Izu, 1977) report that teachers are sensitive to the reliability of the information they receive and can adjust their judgments of students accordingly; teachers differentiate among the data they receive and use them differentially, depending upon the instructional issue they confront.
8. Teachers' perceptions of students' behavior are stable and not much influenced by data when the new information seems to contradict what they have observed (Pedulla, et al., 1977; Morine-Dershimer, 1979; Sorotzkin, et al., 1974; Beggs, et al., 1972).
9. Teachers hold accurate perceptions of what their students' academic achievement test scores and aptitude scores will be (Dusek and O'Connell, 1973; Shavelson, et al., 1977; Airasian, 1980; Brophy and Good, 1970).
10. In contrast to teachers' perceptions of their students' test scores, there is some evidence that teachers' reporting of their students' classroom interpersonal behavior is neither stable nor accurate (Elmore and Beggs, 1972).
11. Teachers seem not to be accurate observers of pupils' academic behavior. Several examples in the literature illustrate teachers' observations of oral reading by their pupils. Regardless of the amount of training or experience, teachers appear to make poor judges of the oral reading behaviors of students (Ladd, 1961; Page and Carlson, 1975; Allington, 1978).
12. When parents are asked to offer information about their children's academic levels, they give information that is inaccurate (Naron, 1977).
13. When specific teacher behaviors are involved, teachers give inaccurate assessments of their own behaviors when checked against the observations of those same behaviors by trained observers. When teachers are asked to assess teaching styles in global terms, they appear to be more accurate (Hook and Rosenshine, 1979).

Is there a relationship between how teachers behave and their students' achievement?

14. Most of the studies that have explored the relationship between teacher behavior and student achievement have concluded that these relationships are either low or non-replicable (Rosenhine, 1970; Rosenhine and Furst, 1973; Heath and Nielson, 1974; Moon, 1971).
15. Those teacher behaviors that have been found to be generalizable and amenable to measurement include teacher-questioning behavior, teacher feedback, and classroom management (when it was positive and did not deal with controlling variables), and teacher-probing behavior related to a student's initial answer to a question (Erlich and Shavelson, 1978).

16. The lack of any high and persistent relationships between teacher behaviors and student achievement appears to be due to three factors: measurement problems, conceptual problems, and the design of the studies addressing this relationship (Erlach and Shavelson, 1978).

How effective is the diagnostic-prescriptive teaching model?

17. The answers to this question are mixed. There is convincing evidence on both sides of the question. Those who have viewed the answer to this question negatively have tended to look at it in the light of special education (Arter and Jenkins, 1979; Hammill, 1972; Ysseldyke, 1973; Larsen and Hammill, 1975).

Those who have addressed this question positively have viewed the diagnostic-prescriptive process in the mainstream of education, including the sciences, social sciences, mathematics, and the like (Yeany, et al., 1979; Goodson and Okey, 1978; Keller, 1968; Frieder, 1970; Bloom, et al., 1971; Hickey and Hoffman, 1973; Block and Burns, 1976).

18. A major argument for the diagnostic-prescriptive teaching model centers on the observation that objectives coupled with diagnostic testing and remediation improve achievement in high school and college-level science courses and elementary school mathematics courses (Yeany, et al., 1979; Long, et al., 1979; Goodson and Okey, 1978; Burrows and Okey, 1975; Swanson and Denton, 1976; Washburne, 1922).
19. Arguments against the diagnostic-prescriptive teaching model include the following: (a) vague definition of psychological and educational abilities hampers effective measurement of these abilities, (b) the reliability of diagnostic tests is not deemed to be high enough, except in a few isolated instances, (c) validity statistics are not deemed high enough for those tests that are used for diagnosis, (d) training specific abilities does not result in improvement nor in higher achievement levels, and (e) diagnosis does not result in improved academic programming (Arter and Jenkins, 1979; Hammill, 1972; Larsen and Hammill, 1975; Newcomer, et al., 1975).

Inservice training of teachers -- a summary. Early in this report to you a constellation of questions were posed concerning the inservice training of teachers in the use of assessment information in the classroom. There was a hope, not fully realized, that a search of the literature would uncover successful patterns of inservice education that could be replicated with confidence. There was the assumption voiced that the inservice education of practitioners in other fields would lead to some suggestions which could be applied with confidence to the preparation of those who needed help with the utilization of assessment information. There was the anticipation that one could identify research in critical incidents of teaching and set about to replicate these in an inservice education program.

It would appear that these expectations will, for the most part, not be realized and will await another sweep of the literature some time in the future. What have we learned to date?

What is appropriate content for an assessment-teaching inservice experience?

1. There is considerable reason to believe that teachers have limited knowledge about fundamental measurement concepts (Goslin, 1967; Farr and Griffin, 1973; Noll, 1955; Olejnik, 1979).
2. An inservice program linking assessment with teaching must surely include decisionmaking models (Shavelson and Borko, 1979; Shavelson, 1978; Rudman, 1977; Farr and Griffin, 1973).
3. An analysis of teacher tasks related to educational measurement indicate seven major tasks that should be included in an inservice program: (1) preparing examinations, (2) administering and scoring tests, (3) test selection, (4) using other nonstandardized evaluative devices and techniques, (5) computing and interpreting statistical data, (6) using test results to plan instruction, and (7) evaluating pupil performance (Rosner, 1961).
4. There is widespread belief among those concerned with teacher education that a rational approach to instructional decisionmaking is a desirable one (Clark and Yinger, 1977; Shields and Neal, 1978; Clemson University Vocational Education Media Center, 1978; Evans, et al., 1975). Most-often-mentioned topics include the framing of performance objectives, development of criterion-referenced measurements, individualizing performance-based education, self-paced program development, and the like.

What have we learned about the structure of inservice experiences?

5. Most inservice experiences (lectures, self-study modules and pamphlets, analysis of video tape incidents) are similar in design and are usually structured for group participation and study (NEA, 1976, 1977a, 1977b, 1977c, 1977d; Deanda, 1975; Capie, 1978; Harrison, 1976).
6. The Active Response Inservice Training Method utilized in special education blends group instruction with tutorial experiences between consultant and teacher and shows some promise as a useful device (Snell, et al., 1979; Snell, et al., 1978; McBride, 1972; Herbert, 1975).
7. Inservice experiences can take a variety of forms and functions throughout the year. It can include the preparation of bulletins germane to testing or instruction, extension courses, invitational workshops, demonstration centers, and to using educational TV for topics of concern to teachers (Durost, 1959).

What impact have inservice education programs had on educational practice?

8. One study of all papers reviewed dealt in a systematic way with the impact of an inservice program on its participants and the subsequent influence on the institute members, on their colleagues, and on school policies (Hastings, et al., 1961).

It was clear that there was a stable and positive influence of institute staff members and programs upon those who participated in the training activities. It was also evident that former students in the institute

had considerable influence over their faculty members when they returned to their schools. But a puzzling outcome was that those institute participants had little if any influence over the direction of opinions or practices that would modify school policies.

What principles or conclusions have we drawn concerning those characteristics of inservice programs that have proven successful?

An inservice program can be most effective if it is:

1. predicated on a long-term plan as opposed to a single-session meeting;
2. one that utilizes personal contact between the trainer and those being trained;
3. a program that utilizes onsite contact between teachers and trainer, rather than a program conducted away from the site where a reasonable opportunity exists for application of skills and concepts learned;
4. based upon collaboration between teachers, administrators, and those specialists charged with conducting the inservice program;
5. structured so that there is an opportunity for the real world of the classroom to intermesh with those materials, concepts, and skills to be learned;
6. predicated on a followup program that utilizes local staff and specialists who may be external to the school and/or system;
7. based upon a plan that utilizes demonstration, supervised trials, and feedback;
8. one that utilizes the opportunity for teachers to relate to other teachers who may serve as consultants to their colleagues.

What Is and Has Been Available to Teachers to Aid Them in Interpreting and Using Assessment Information for Instruction Decisionmaking? A serious obstacle to the researcher interested in discovering the relationship of assessment to teaching is the limitation imposed by ERIC policies restricting the amount of copyrighted material catalogued in a given time period. Almost all reviews of the literature dealing with testing and teaching have overlooked the availability of materials designed to help teachers and administrators interpret and use tests to implement instructional planning. Some (Yeh, 1978; Kohn, 1975) have complained of the lack of cooperation of test publishers in supplying the information and materials to them. The staff of this project did not experience the same difficulty. Five publishers who produce the majority of tests designed for use in grades kindergarten through 12th grade were asked to supply copies of all interpretive manuals and other devices that accompany the tests they published for use in the elementary, middle, and high schools. With no exception, all did so. In addition, the United States Department of Defense (1979) supplied the reviewers with a filmstrip and cassette designed to orient teachers in the DODDS system on the proper interpretation of tests used that employ a latent trait construction and reporting mode; the National Education Association submitted a sample of the materials they produce, as did the Michigan Department of Education.

Time limitations as specified by contract between NIE-DODDS and Michigan State University prevented an extensive analysis of interpretive materials available to teachers by the various State departments of education. One such example, however, was reported in the review published by our Institute for Research on Teaching.

The format of this section of that review differs from preceding sections. The review of these materials in no way suggests that these are or will continue to be the extent of interpretive materials available from these sources. Nor are these sources the only ones publishing interpretive aids for the practitioner.

That section of the review was included to indicate that help is available to the teachers and administrators and has been available for several decades, in one form or another.

Time constraints, and your collective patience, prohibit any detailed review of those materials. However, it would be useful to share with you the process developed for the review of the submitted materials, the elements we agreed to review, and the criteria we employed in the selection of materials to be described.

Each staff member selected a publisher to review and worked independently on the materials submitted. The staff agreed upon a standard format to be employed wherever possible so that comparisons between materials could be made by the readers of the review. An outline of all of the materials described for each source preceded the detailed descriptions of interpretive aids. To avoid any subsequent misunderstandings, the project director did not contribute any reviews of the materials submitted by any of the five commercial publishers. His coauthorship of a major achievement test series, while an asset to the team, dictated some prudence.

The staff agreed before describing the various materials to avoid as much as possible any significant value judgments concerning any of the materials chosen for review. While the reviews are not free of all value calls, they do reflect a judicious attempt to be as neutral as possible. The major purpose for the reviews was not evaluative, but informative: what was available to teachers and others?

Each description included identification of authorship, titles of manuals reviewed, the identification of levels covered by the tests and interpretive devices, the publisher, a description of the types of scores available, a description of how the results were reported, and aids available for the interpretation and use of the results.

Materials were included if they were specifically addressed to the interpretation and use of tests in teaching and in curriculum planning. Some materials, while dealing with the issues surrounding the use of tests, were not deemed to be of direct help to teachers who were faced with the specific question of "How can I use assessment information in my classroom?" Some of the materials submitted really addressed another question, viz., "Why shouldn't I use standardized tests in my classroom?" While this question is a legitimate one, it lay outside the purview of this project.

What Was Learned From the Literature Review That Was New, Different, or a Surprise? Probably the most disturbing element in the review was the work reported by Stetz and Beck in a series of studies reported during the years 1977 through 1980. These four studies addressed the attitudes of those who either used

standardized achievement tests or were in some way involved in the dissemination of these results. The prevailing message that permeated these reports was that much of the material written about teacher use of tests was written by those who were not the actual users of tests. While this may not be a serious problem, it does raise the question of accuracy of the perceptions of teachers and others among those who write about them.

The latest in this series of studies by Beck and Stetz was reported at the 1980 meetings of the National Council on Measurement in Education (NCME) and the American Educational Research Association. A random sample of 500 NCME members was drawn from the NCME membership list. The responses were compared to the data previously collected among teachers and students who had taken part in a standardization of the 1978 edition of the Metropolitan Achievement Tests. The following selected highlights of the results of this investigation are particularly germane to the issue of whether those who write about teachers' perceptions write with some accuracy:

... Amount of Standardized Testing That Takes Place

Almost two-thirds of NCME members think teachers would reply, "too much," while less than one-fourth actually did so.

... Teachers' Personal Use of Standardized Achievement Test Results

Half of each group indicated "some" use made, while just under 10 percent of each group indicated "considerable" teacher use. On this point, then, NCME members perceived teachers' behaviors accurately.

... Teachers' Attitudes Towards Standardized Achievement Tests

Teachers report that they are more interested in, calm about, comfortable with, and knowledgeable about standardized tests than NCME members thought they would be.

... Teachers' Desire for Further Measurement Training

Fewer teachers than expected desire additional training in test interpretation. (60 percent of the teachers versus 76 percent of NCME members).

... Teachers' Desires for Increased Use of Criterion-Referenced Tests

Seventy-four percent of NCME members thought that teachers wanted to increase their use of criterion-referenced tests; only 35 percent of the teachers actually wanted to do so.

... Purposes for Using Standardized Test Results

Teachers reacted more positively to using test data for measuring "growth" than expected by NCME members (77 percent vs. 54 percent), planning instruction (65 percent vs. 43 percent), evaluating teaching methods (26 percent vs. 15 percent), and evaluating teacher performance (19 percent vs. 1 percent).

The authors drew attention to the significance of their findings by concluding the following:

NCME is a professional association of measurement specialists that is, by design, practitioner-oriented. As such, it is important that its members be sensitive to the attitudes of test users towards assessment devices. This study highlights areas in which perceptions of specialists about user attitudes vary markedly from reality. It is important to monitor the relationship between perceived and actual attitudes if NCME is to be successful in assisting non-specialist test users in making better use of educational assessment.... (Beck and Stetz, 1981, p. 4.)

One could add to the authors' concerns that, since a significant amount of the prescriptive and descriptive contributions to the literature come from specialists who do not normally teach in K-12 classrooms, the accuracy of their conclusions needs to reflect the accuracy of their insights of teachers' perceptions of test use and interpretations.

A second, and related, issue had to do with the differentiated attitudes of teachers toward testing and test use. Too much of what has been written about either teachers' apathies toward testing, or their support of testing has missed a crucial element. Familiarity with testing, teaching experience, and children's performance on tests all have a bearing upon the attitudes that are held about test use.

Others have addressed portions of the issues concerning Beck and Stetz (Kirkland, 1971; Brim, et al., 1965; Brim, 1965; Brim, et al., 1964), but perhaps the most often cited has been the work of Goslin (1965, 1967). Goslin's book, Teachers and Testing (1967) is based upon two previous studies developed with the support of the Russell Sage Foundation (Goslin, et al., 1965; Goslin, 1965).

One of the findings reflected in Goslin's works had to do with the impact of measurement knowledge on teachers' attitudes toward test use. He reported that those teachers who had more training in measurement concepts and experience in the administration and interpretation of tests generally tended to make the most of them. But, unfortunately, both elementary and secondary teachers had little if any formal training in educational measurement (only 40 percent of elementary and secondary teachers had as little as one class in testing) (Goslin, 1965).

Jennie Yeh and her colleagues at the Center for the Study of Evaluation (CSE) undertook a comprehensive national survey of test use in the schools (Yeh, 1978). She contacted 51 chief state school officers, directors of instructional research in 64 school districts with enrollments of 50,000 or more students, and 11 publishers of tests used within the United States. She also used 20 schools located within five California school districts to conduct interviews with teachers and administrators.

Findings that relate to this section of the review indicate the following:

...Teachers felt that the direction of their testing programs were, in some measure, influenced by the concerns shown for test results by parents of pupils enrolled in their classes.

...Student performance influenced teachers' attitudes toward measurement practices. Staffs of schools whose children did poorly on educational measures exhibited the least inclination to use test results.

...Conversely, in those schools in which pupil performance on tests was high, teachers tended to view tests as useful.

...Generally, the more knowledgeable one was about testing, the more likely was one to support the usefulness of all types of educational measurement.

...When teachers chose commercially prepared tests, they considered most important, clear format, similarity to class material, and accurate prediction of achievement. These were, incidentally, the same reasons that teachers gave for developing their own tests.

...Teachers considered the time spent in required testing vs. the benefits that accrued from such testing as falling between "about right" and "slightly too much."

...The cost benefit differences varied significantly depending upon the teachers' experience. Teachers with less than 16 years' experience were more likely to respond that too much time was spent in relation to the benefits received than were those teachers who had more than 16 years of experience.

...Teachers with more teaching experience were more likely to use test results from norm-referenced tests than were less-experienced teachers (Yeh, 1978).

Teachers' test use patterns. A troublesome aspect in this area is the paucity of descriptive material compared to the abundance of prescriptive articles, essays, and the like dealing with the specifics of how teachers used test results in their classroom. When coupled with the information supplied by Beck and Stetz (in press) concerning the relatively inaccurate perceptions of measurement specialists who write about teacher testing behavior, positive conclusions about how teachers use tests can be only fragile speculations at best. However, much of what has been prescriptively written is logical and in accord with what is recognized as sound psychometric practice. Much of what has been prescribed has been written by measurement specialists with well-deserved reputations. The concern voiced here is not with good prescription but with the lack of accurate and abundant descriptions of test practice.

Finally, our review of the materials presently available to teachers from various sources leads us to the observation that these materials are good and are helpful in understanding how to interpret the derived scores in which the results are counted. But a weakness in much of the material presently available is a lack of focus on teaching acts and on the inability of test users and test publishers to disseminate them widely to those who need them the most. And that observation brings me to a last description of our first year of work in this assessment project.

The Data Box. Our team has developed and is continuing to develop a set of instructional materials to aid teachers in the understanding and use of assessment information of all kinds, formal and informal alike. Our work has focused on two major dimensions: the fundamental measurement concepts that we believe are essential for intelligent use of information, and those critical events in teaching

that occur throughout a school year. Our mechanism for capturing these two dimensions has been the use of vignettes that describe a teaching situation and ask for the identification and interpretation of documents that would yield appropriate information and/or actions that could be taken.

The system has been packaged appropriately and consists of vignettes, documents, and an interpretive manual. We are now in the process of evaluating the materials with practitioners to determine the degree to which they reflect the reality of the classroom and the types of documents teachers normally encounter. We hope to evaluate the entire system along with several different approaches to disseminating the materials and training teachers in the use of them. What do we hope to see at the end of that evaluation period? We hope to see teachers who are more secure in the use of information to back their intuitive judgments, who can use all types of formal and informal data with a degree of confidence, and who have an increased ability to make judgments that reflect the information they have received. This would indicate to us a capacity for understanding these basic measurement concepts I speak about. We would hope that teachers would be (1) more skillful and less prone to over- and under-interpret data, (2) would recognize the parameters of the data being considered, and (3) would use these data to arrive at those professional judgments that only teachers can make. This is a tall order, we know. We are anxious to discover what our efforts have wrought.

7. CONFERENCE SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Lawrence M. Rudner, National Institute of Education

It is inaccurate to view testing as a minor part of the instructional process which occurs only occasionally and on an isolated basis. Students are assessed for strengths and weaknesses daily through the use of observation, interaction, homework, instructional activities, textbook quizzes, and workbook questions; it is an integral part of instruction. The results of the assessment are used for placement and selection of students, tailoring individual and classroom instruction, selection of curriculum, and reporting to students, parents, and school administrators.

Standardized testing, including norm-referenced achievement tests, is perhaps the most visible and controversial form of student assessment. Two hundred million achievement test forms are used each year, with an annual cost of at least \$24 million to public elementary and secondary school students. While standardized tests are perhaps the most studied form of assessment, questions remain as to the recommended amount and the nature of the use of these measures in the classroom.

The National Institute of Education (NIE) invitational conference on test use brought together a variety of researchers who had investigated the use of assessment information in the classroom. James Ward, of the American Federation of Teachers (AFT), surveyed a sample of 200 AFT members to collect information needed for the development of training materials addressing the use of standardized achievement tests. Leslie Salmon-Cox, of the Learning Research and Development Center of the University of Pittsburgh, conducted detailed interviews with 35 teachers in three elementary schools as part of a study funded by the Carnegie Corporation. Jennie Yeh, of the UCLA Center for the Study of Evaluation, has been conducting a program of research on the usage of tests in schools. Her paper focused on the results of a survey administered to 260 teachers in a stratified sample of California schools.

Michael Beck, of the Psychological Corporation, summarized the results of three large-scale surveys of teachers, students, and measurement professionals. Herbert Rudman, of Michigan State University, reported the results of a literature review conducted in preparation for the development of teacher training materials for linking multiple forms of assessment to instruction. Mary Kennedy, of the Huron Institute, discussed test use by teachers and program directors interviewed as part of a study of school districts cited for their exemplary use of test and education data.

Through the use of different methods and with different purposes, these papers presented a fairly uniform picture of responsible test use in American schools. This picture does not reflect either the pro- or anti-testing stances in the literature and does not reflect the perceptions of measurement specialists.

Beck compared the results of a survey of a random sample of 284 members of the National Council on Measurement in Education -- a practitioner-oriented group of measurement specialists -- to the results of a survey of a stratified random sample of 3,500 classroom teachers participating in the norming of a standardized test. He concluded that measurement specialists generally do not fully understand the opinions about standardized testing held by users of such instruments nor how the

instruments are used. A majority of measurement specialists erroneously thought that a majority of teachers would feel that too much testing was conducted in their system, that the primary use of test results was to report to parents, that standardized tests are useful for comparing students with national peer groups, and that moratoriums on standardized intelligence tests and state-mandated tests are desirable.

Contrary to the apparent expectations of the measurement specialists, teachers indicated that standardized testing plays a relatively minor role in instruction and assessment (Yeh; Beck; Salmon-Cox). Teachers use tests to supplement or confirm the results of other assessments and do not feel test results should be used in isolation (Salmon-Cox; Kennedy).

As part of the everyday teaching experience, teachers diagnose student strengths and weaknesses and desire this type of information (Salmon-Cox). The primary use of standardized tests, to diagnose individual students, reflects this need. However, standardized testing is neither designed nor fully able to meet this need. Further, the nature of these reported diagnostic uses is not clear from the research. Teachers view standardized tests as limited by the amount of detail provided, by their lack of prescriptive information, by the response time for results, and by the lack of curriculum match (Ward). They tend to prefer tests that are built into their textbooks or are closely matched to their curriculum (Kennedy). Some of these limitations may reflect both the low percentage (10-20 percent) of schools requesting supplemental student item analysis information and the delays in sending answer forms to test publishers.

Standard tests, then, are used occasionally, and only as one form of assessment input into the instructional process. They are used primarily for individual assessments of strengths and weaknesses rather than for evaluating program or curricular content. Teachers show a responsible attitude toward standardized tests by using them in conjunction with other multiple forms of assessment. This use of tests is a restriction in the range of possible uses of tests. Uses not related to instruction, such as comparing students with national norms, reporting to newspapers, pondering instruction, and evaluating curriculum are viewed as the least useful purposes of testing (Beck, Ward, Salmon-Cox).

Teacher tests, observation, curriculum tests, and informal means, rather than standardized achievement tests, are the primary means used to assess students throughout the school year (Yeh; Salmon-Cox; Kennedy). The major reasons cited are that they are more suitable for students, they better assess the instruction given, and they provide more diagnostic information (Salmon-Cox; Ward; Yeh). In light of their prominence and function, these are areas that appear to be insufficiently stressed by measurement specialists.

The emphasis on local assessment does not appear to be due to teacher discomfort with standardized tests. Teachers are more interested in, and more calm, comfortable, and knowledgeable about standardized tests than measurement specialists generally give the credit for (Beck). Teachers feel they know how to use tests for instructional planning and are comfortable with standardized scores (Ward, Rudman).

Teachers are most interested in diagnostic testing and generally stress the importance of relating it to that purpose. Teachers feel it is important to have expertise in selecting a test and in judging the appropriateness of tests (Ward). They use sound criteria -- format, content validity, predictive validity, and ease

of administration (Yeh) -- to evaluate tests. Even though they may not have expertise in interpreting grade equivalent scores or percentile scores, teachers apparently have the skills necessary for the limited uses they make of tests. It has been noted, however, that teachers with the most training in measurement concepts generally make the most use of test results.

The greater emphasis on teacher-made tests appears to be reflected in student attitudes. Students are generally more anxious before taking a teacher-developed test than they are before taking a standardized test (Beck), and they feel teacher-developed tests are harder (Beck). One can conjecture that the low frequency, the low emphasis, and the fact that test results are not frequently reported basically neutralize student attitudes toward standardized tests (Beck). This neutral attitude should not be interpreted as a lack of interest.

Conclusions and Recommendations

The NIE Invitational Conference on Test Use was held in 1980, only 2 years after a larger Conference on Achievement Testing and Basic Skills (Tomlinson and Treacy, 1979) was convened by HEW Secretary Joseph A. Califano. The additional information provided by the test-use conference reinforced two major conclusions:

- Education goals should drive the agenda of testing and evaluation, and not the other way around.
- Instead of asking how we can make better tests, we should ask how we can make better use of tests, how tests can be effective means of diagnosis, and how those diagnoses can in turn be translated into educational intervention, remediation, and compensatory efforts.

Additionally, the NIE Conference on Test Use arrived at the following conclusions:

- Teachers use tests in conjunction with other sources of information to assess individual students.
- Teacher-developed and curriculum-embedded tests are the primary means of assessing students.
- Measurement specialists generally are not in tune with how tests are used in the classroom.
- Teachers desire and use diagnostic information.

While there is an increase in the amount of practical, school-oriented testing research being conducted, there are numerous unanswered questions. As a result of the papers presented and conference discussions, the following research questions were identified:

- What are alternatives to testing for students' diagnoses and assessment?
- How can one create a climate fostering better test use?
- How much do parents know about testing? What are their attitudes? What are the best means of communicating student achievement?

- What are student attitudes and reactions to testing?
- Do teacher-developed tests cover a range of cognitive skills levels? Are they valid, reliable, and effective? What can be done to improve teacher-developed tests?
- Do curriculum-embedded tests cover a sufficient range of cognitive skills? Are they valid, reliable, and effective? What can be done to improve them?
- What are characteristics of useful diagnostic tests?
- How can tests be used to provide better diagnostic information about students, groups of students, and curricula?
- How can teachers develop better diagnostic instruments? How can test publishers develop better diagnostic instruments?
- What are alternative ways of reporting test results to students? Parents? The press?
- How can local districts better analyze test results?
- How do teachers use standardized tests diagnostically?

REFERENCES AND BIBLIOGRAPHY ON TEST USE*

- Airasian, P.W., "Uses and Misuses of Standardized Tests." Measurement in Education 10 (1980).
- Airasian, P.W. and Madaus, G.F., "Criterion-Referenced Testing in the Classroom." Measurement in Education 3 (1972): 1-8.
- Allington, R.L., "Teacher Ability in Recording Oral Reading Performance." Academic Therapy 14 (November 1978): 187-192.
- American Federation of Teachers, Policy Statement, 1976 AFT Convention. Further discussion in "What Do Teachers Think About Tests and Testing?" American Educator, Winter, 1977.
- Ammons, R.B., Knowledge of Performance, Survey of Literature, Some Possible Applications, and Suggested Experimentation. WADC Technical Report 54-14, Wright Development Center, 1954.
- Anastasi, A., Psychological Testing. 3rd ed. New York: MacMillan 1968.
- Anderson, R.C., Kulhary, R.M., and Andre, T., "Feedback Procedures in Programmed Instruction." Journal of Educational Psychology 62 (1971): 148-156.
- Annett, J., The Role of Knowledge of Results in Learning: A Survey. U.S. NAVTRADEVCECEN Technical Document Report 342-343, 1961.
- Arter, J.A. and Jenkins, J.R., "Examining the Benefits and Prevalence of Modality Considerations in Special Education." Journal of Special Education 11 (1977): 281-298.
- Arter, J.A. and Jenkins, J.R., "Differential Diagnosis -- Prescriptive Teaching: A Critical Appraisal." Review of Educational Research 49 (Fall 1979): 517-556.
- Barnard, J.W., Zimbardo, P.G., and Sarason, S.B., "Teachers' Ratings of Student Personality Traits as They Relate to IQ and Social Desirability." Journal of Educational Psychology 59 (1968): 128-132.
- Barr, R., "How Children Are Taught to Read: Grouping and Pacing." School Review 83 (1975): 479-498.
- Beatty, L., Madden, R., Gardner, E., and Karlsen, B., Stanford Diagnostic Mathematics Tests. New York: Psychological Corporation, 1976.
- Beatty, L.S., Madden, R., Gardner, E.F., and Karlsen, B., Stanford Diagnostic Mathematics Test (1 to community college). "Manual for Administering and Interpreting." New York: Harcourt, Brace, Jovanovich, 1976.

*The vast majority of these references were compiled by Herbert C. Rudman, Janet L. Kelly, Donna S. Wanous, William A. Mehrens, Christopher M. Clark and Andrew Porter, Institute for Research on Teaching, Michigan State University.

- Beck, M.D. and Stetz, F.P., "Teacher Opinions of Standardized Test Use and Usefulness." Paper presented at annual meeting of American Educational Research Association, San Francisco, April 1979.
- Beck, M.D. and Stetz, F.P., "Standardized Testing as Viewed by Test Specialists and Users." Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, 1980.
- Beggs, D., Mayer, G.R., and Lewis, E.L., "The Effects of Various Techniques of Interpreting Test Results on Teacher Perception and Pupil Achievement." Measurement and Evaluation in Guidance 5 (April 1972): 290-297.
- Bennett, G.K., Seashore, H.G., and Wesman, A.G., Differential Aptitude Tests (grades 8-11). "Fifth Edition Manual," 1973, and "Counseling from Profiles," 1977. New York: The Psychological Corporation.
- Bhaerman, R., "What Do Teachers Think About Tests and Testing?" American Educator (Winter, 1977).
- Block, J.H. and Burns, R.B., "Mastery Learning." In Review of Research in Education. L. Shulman, editor. Itasca, Illinois: Peacock Publishers, 1976.
- Bloom, B.S., "Learning for Mastery." Evaluation Comment 1 (1968): 273-281.
- Bloom, B., Hastings, J., and Madaus, G., Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill Book Co., 1971.
- Borg, W., Kallenback, W., Kelley, M., and Langer, P., A Microteaching Approach to Teaching Education. Beverly Hills, California: MacMillan Educational Services, 1970.
- Borko, H., "Factors Contributing to Teachers' Preinstructional Decisions about Classroom Management and Long-Term Objectives." Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.
- Boyd, J., Jacobsen, K., McKenna, B.H., Stake, R.E., and Yashinsky, J., A Study of Testing Practices in the Royal Oak (Michigan) Public Schools. Michigan: Royal Oak Michigan School District, 1975.
- Brent, G., Glassboro State College - 12th Cycle Teacher Corps Program. Washington, D.C.: Teacher Corps, Office of Education, 1976.
- Brim, O.G., Jr., Goslin, D.A., Glass, D.C., and Goldberg, I., The Use of Standardized Ability Tests in American Secondary Schools and Their Impact on Students, Teachers and Administrators. Technical Report No. 3. New York: Russell Sage Foundation, 1964.
- Brim, O.G., Jr., "American Attitudes Toward Intelligence Tests." American Psychologist 20 (1965): 124-130.
- Brim, O.G., Jr., Neulinger, J., and Glass, D.C., Experience and Attitudes of American Adults Concerning Standardized Intelligence Tests. Technical Report No. 1. New York: Russell Sage Foundation, 1965.

- Brophy, J.E. and Good, T.L., "Teachers' Communication of Differential Expectations for Children's Classroom Performance: Some Behavioral Data." Journal of Educational Psychology 61 (1970): 365-374.
- Brown, L. and York, R., "Developing Programs for Severely Handicapped Students: Teacher Training and Classroom Instruction." Focus on Exceptional Children 6 (1974): 1-12.
- Buckley, P.K. and Cooper, J.M., "An Ethnographic Study of an Elementary School Teacher's Establishment and Maintenance of Group Norms." Paper presented at annual meeting of American Educational Research Association, Toronto, March 1978.
- Buros, O.K., editor, Seventh Mental Measurements Yearbook, Vol. II. Highland Park, New Jersey: The Gryphon Press, 1972.
- Burrows, C.K. and Okey, J.R., "The Effects of a Mastery Learning Strategy." Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C., April 1975.
- Calfee, R. and Juel, C., "How Theory and Research on Reading Assessment Can Serve Decision-Makers." Paper presented at Minnesota Perspectives on Literacy Conference, June 1977. (ED 146539)
- Capie, W., et al., Teacher Performance Assessment Instruments: Plans for Practice Rating. Atlanta: Georgia State University, 1978. (ED 161956).
- Clark, C.M. and Yinger, R.J., "Research on Teacher Thinking." Curriculum Inquiry 7 (1977): 279-304.
- Clark, C.M., From a memorandum dated August 17, 1979, pertaining to current research related to this proposed developmental project.
- Clemson University Vocational Education Media Center. Implementing Performance-Based Vocational Education Utilizing V-Tecs Catalogues. Washington, D.C.: Office of Education, 1978. (ED 159382).
- Davidson, H.H. and Lang, G., "Children's Perception of Their Teacher's Feelings Toward Them Related to Self-Perception, School Achievement and Behavior." Journal of Experimental Education 29 (1960): 107-118.
- Dasho, S., "A Communications Approach to the Negotiation of Powers in the Classroom." Paper presented at annual meeting of American Educational Research Association, Toronto, March 1978.
- Deanda, N., Helping Children Develop Healthy Self-Concepts. Washington, D.C.: National Institute of Education, 1975. (ED 129455).
- Durost, W.N., "Problems in In-Service Training of Teachers in the Use of Measurement and Evaluation Techniques." National Council on Measurement in Education, 16th Yearbook, 1959.

- Dusek, J.B. and O'Connell, E.J., "Teacher Expectancy Effects on the Achievement Test Performance of Elementary School Children." Journal of Educational Psychology 65 (1973): 371-377.
- Dusek, J.B., "Do Teachers Bias Children's Learning?" Review of Educational Research 45 (1975): 661-684.
- Ebel, R.L., "Educational Tests: Valid? Biased? Useful?" Phi Delta Kappan 57 (October 1975): 83-86.
- EDC News. Standardized Testing in America. Spring 1977, Issue No. 9.
- Edgar, E., Baker, S., Harper, C., Swift, P., and Melseth, S., "All Individualized Inservice Training Program for Teachers of the Mentally Retarded." Education and Training of the Mentally Retarded 11 (1972): 77-80.
- Educational Testing Service. CIRCUS (preschool to fall grade 3). "Manual and Technical Report." Menlo Park, California: Addison-Wesley, 1978.
- Educational Testing Service. CIRCUS (preprimary to grade 3). "After the Circus: Guide to Interpretation of Scores and Teacher's Idea Book for Follow-up Instruction in Preprimary Classes" and "Circus Slide Show Presentation." Menlo Park, California: Addison-Wesley, 1977.
- Educational Testing Service. CIRCUS (preschool to fall first). "Circus A and B User's Guide: Activities Inventory" and "Educational Environment Questionnaire." Menlo Park, California: Addison-Wesley, 1975 and 1976.
- Educational Testing Service and National Consortium of School Districts. Basic Skills Assessment Program (8-12). "Technical Manual for Basic Skills Assessment Program/Reading, Mathematics, Writing." Menlo Park, California: Addison-Wesley, 1978.
- Educational Testing Service. CIRCUS (spring 1 to fall 2, spring 2 to fall 3). "Circus C and D User's Guide" and "Basic Assessment/Reading, Mathematics, Listening, Writing." Menlo Park, California: Addison-Wesley, 1979.
- Educational Testing Service. CIRCUS (preschool to fall 3). "User's Guides for A, B, C, and D." Menlo Park, California: Addison-Wesley, 1975-1979.
- Educational Testing Service. Sequential Tests of Educational Progress (spring 3 to spring 12). "Step III: Manual and Technical Report" (preliminary), "Step III: Sample Reports." Menlo Park, California: Addison-Wesley, 1979.
- Elmore, P. and Beggs, D.L., "Stability of Teacher Ratings of Pupil Behavior in a Classroom Setting." Paper presented at a meeting of the American Personnel and Guidance Association, March 1972. (ED 066685)
- Evans, H., et al., A Manual for the Implementation in Local School Districts of the Project APT Inservice Training Course for Teachers. Washington, D.C.: Bureau of Elementary and Secondary Education, 1975. (ED 148050)

- Parr, R. and Griffin, M., "Measurement Gaps in Teacher Education." Journal of Research and Development in Teacher Education 7 (1973): 19-28.
- Feldhusen, J.F., "Student Perceptions of Frequent Quizzes and Post-Mortem Discussion of Tests." Paper presented at annual meeting of the National Council on Measurement in Education, Chicago, February 1964.
- Feshbach, N.D., "Student Teacher Preferences for Elementary School Pupils Varying in Personality Characteristics." Journal of Educational Psychology 60 (1969): 126-132.
- Finn, J., "Expectations and the Educational Environment." Review of Educational Research 42 (1972): 387-410.
- Flanders, N.A., Teacher Influence, Pupil Attitudes and Achievement, Cooperative Research Monograph No. 12. Washington, D.C.: U.S. Department of Health, Education and Welfare, Office of Education, 1965.
- Fleming, E. and Anttonen, R., "Teacher Expectancy or My Fair Lady." American Educational Research Journal 8 (1971): 241-252.
- Fleming, M., "Standardized Tests Revisited." School Counselor 19 (November 1971): 71-72.
- Freeman, D., Kuhs, T., Knappen, L., and Porter, A., "A Close Look at Standardized Tests." Arithmetic Teacher (in press).
- Frieder, S., "Motivation: Least Developed of Teacher Roles." Educational Technology 10 (1970): 28-36.
- Garcia-Quintana, R.A. and Johnson, L.M., "Using Norm-Referenced Data to Set Standards for a Minimum Competency Program in the State of South Carolina: A Feasibility Study." Presented at annual meeting of the American Educational Research Association, San Francisco, April 1979.
- Glass, G.V., "Primary, Secondary and Meta-Analysis of Research." Paper presented at annual meeting of the American Educational Research Association, San Francisco, April 1976.
- Goodman, Y., "Using Children's Reading Miscues for New Teaching Strategies." Reading Teacher 23 (1970): 455-459.
- Goodson, M.L. and Okey, J.R., "The Effects of Diagnostic Tests and Help Sessions on College Science Achievement." Journal of College Science Teaching 8 (November 1978): 89-90.
- Gorth, W.P. and Perkins, M.R., A Study of Minimum Competency Testing Programs: Final Summary and Analysis Report. Amherst, Massachusetts: National Evaluation Systems, Inc., 1979. (ED 185124)
- Goslin, D.A., Epstein, R., and Hallock, B.A., The Use of Standardized Tests in Elementary Schools. Second Technical Report. New York: Russell Sage Foundation, 1965.

- Goslin, D.A., The Use of Standardized Tests in American Secondary Schools and Their Impact on Students, Teachers, and Administrators. New York: Russell Sage Foundation, 1965.
- Goslin, D.A., Teachers and Testing. New York: Russell Sage Foundation, 1967.
- Gronlund, N.E., Preparing Criterion-Referenced Tests for Classroom Instruction. New York: MacMillan, 1974.
- Hammill, D., "Training Visual Perception Processes." Journal of Learning Disabilities 5 (1972): 39-44.
- Harrison, S., et al., Information for Inservice Training Development: A Collection from Field Efforts of the Rocky Mountain Regional Resource Center. Washington, D.C.: Bureau of Education for the Handicapped, 1976. (ED 238025)
- Hastings, J.T., Runkel, P., and Damrin, D., Effects on Use of Tests by Teachers Trained in a Summer Institute. Urbana, Illinois: Bureau of Educational Research, 1961. (ED 002925)
- Heath, R.W. and Nielson, M.A., "Research Basis for Performance-Based Teacher Education." Review of Educational Research 44 (1974): 463-484.
- Herbert, B., The Active Response Inservice Training Method as Method for Use with Teachers and Aides in Public School Programs for Moderate to Severely Retarded Students. Doctoral dissertation, Michigan State University, 1975.
- Hermanowicz, H., Teacher Education: An Exploration of Efforts and Proposals for Its Improvement. Springfield, Illinois: Illinois Board of Higher Education, 1968. (ED 030589)
- Hickey, M.E. and Hoffman, D.H., "Diagnosis and Prescription in Education." Educational Technology 13 (October 1973): 35-37.
- Hieronymus, A.N., Lindquist, E.F., Hoover, H.D., Oberley, K.R., Lewis, T.B., Neckere, E.D., Santos, B.N., Peterson, J.J., Whitney, M.A., Renk, R.A., Strayer, F.H., Flake, B.S., Iowa Test of Basic Skills (9-14). "Teacher's Guide for Administration, Interpretation and Use". Iowa City: Riverside Press, 1978.
- Hoepfner, R., Bradley, P.A., and Doherty, W.J., "National Priorities for Elementary Education." CSE Monograph Series in Evaluation, 2 (1973).
- Hook, C. and Rosenshine, B.V., "Accuracy of Teacher Reports of Their Classroom Behavior." Review of Educational Research 49 (Winter 1979): 1-12.
- Houts, P.L., "Standardized Testing in America, II." The National Elementary Principal, 6, (1954): 2-3.
- Jones, F.C., "Mastery Learning and Geography: Effect upon Achievement, Retention and Time-to-Completion." Paper presented at annual meeting of the college and university faculty of the National Council for the Social Studies, November 1974. (ED 099280)

- Joyce, B., Howey, K., and Yarger, S., "Special Education and Teacher Education." Special Education and Teacher Education 1 (1977): 67-75.
- Karlsen, B., Madden, R., and Gardner, E., Stanford Diagnostic Reading Tests. New York: The Psychological Corporation, 1977.
- Karlsen, B., Madden, R., and Gardner, E.F., Stanford Diagnostic Reading Test (1 to community college). "Manual for Administering and Interpreting," 1976, and "Handbook of Instructional Techniques and Materials," 1978. New York: Harcourt, Brace, Jovanovich.
- Kellaghan, T., Madaus, G.F., and Airasian, P.W., The Effects of Standardized Testing. Dublin, Ireland: Educational Research Center, St. Patrick's College, 1980.
- Keller, F., "...Goodbye Teacher." Journal of Applied Behavioral Analysis 1 (1968): 78-89.
- Kirkland, M.C., "The Effects of Tests on Students and Schools." Review of Educational Research 41 (1971): 303-350.
- Kohn, S.D., "The Numbers Game: How the Testing Industry Operates." National Elementary Principal 54 (July/August 1975): 11-23.
- Krug, E.A., Curriculum Planning. New York: Harper and Brothers, 1950.
- Ladd, E., A Comparison of Two Types of Training With Reference to Developing Skill in Diagnostic Oral Reading Testing. Unpublished doctoral dissertation, Florida State University, 1961.
- Larsen, S.C. and Hammill, D.D., "The Relationship of Selected Visual-Perceptual Abilities to School Learning." Journal of Special Education 9 (1975): 281-291.
- Lerner, B., "Equal Protection and External Screening: Davis, DeFunis, and Bakke." In Proceedings of the 1977 ETS Invitational Conference. Princeton, New Jersey: Educational Testing Service, 1978.
- Lerner, E., "Tests and Standards Today: Attacks, Counterattacks, and Responses." New Directions for Testing and Measurement, Roger T. Lennon (ed.). San Francisco: Jossey-Bass, Inc., 1979.
- Long, J.C., et al., "The Effects of a Diagnostic-Prescriptive Teaching Strategy on the Achievement and Attitudes of High School Biology Students." Paper presented at the annual meeting of the National Association for Research in Science Teaching, Atlanta, March 1979. (ED 171578)
- Lortie, D.C., School-Teacher -- A Sociological Study. Chicago: The University of Chicago Press, 1972.
- Ludwig, D.J. and Maehr, M.J., "Changes in Self-Concept and Stated Behavioral References." Child Development 38 (1967): 453-469.

- MacGinitie, W.H., Kamons, J., Kowalski, R.L., MacGinitie, R.K., Mackay, T., Gates-MacGinitie Reading Tests (Second Edition). "Teacher's Manual, Level D., Forms 1, 2, and 3." Boston: Houghton-Mifflin, 1978.
- Madden, R., Gardner, E.F., Rudman, H.C., Karlsen, B., and Merwin, J.C., Stanford Achievement Test (1.5 to 9.5). "Teacher's Directions for Administering," "Norms Booklet," "Teacher's Guide for Interpreting," "Practice Booklet and Directions for Administering," "Stanford Index of Instructional Objectives," "Administrator's Guide for Interpreting," "HBJ Scoring Service Reports for Stanford/TASK." New York: Harcourt, Brace, Jovanovich, 1973.
- Mann, L., "Psychometric Phrenology and the New Faculty Psychology: The Case Against Ability Assessment and Training." Journal of Special Education 5 (1971): 3-65.
- Marso, R.N., "Classroom Testing Procedures, Test Anxiety and Achievement." Journal of Experimental Education 38 (Spring 1970): 54-58.
- McBride, H.J., The Differential Effectiveness of Two Methods of Training Institutional Attendants in the Techniques of Behavior Modification. Doctoral dissertation, Michigan State University, 1972.
- McKenna, B., "What's Wrong With Standardized Testing?" Today's Education (March/April 1977): 36.
- Mehrens, W.A. and Ebel, R.L., "Some Comments on Criterion-Referenced and Norm-Referenced Tests." Measurement in Education 10 (Winter 1979): 1-8.
- Mills, C.N. and Hambleton, R.K., "Issues and Methods of Reporting Criterion-Referenced Test Scores." Measurement in Education, 1981.
- Mills, R. and Bryan, M.M., Testing...Grouping: The New Segregation in Southern Schools. Atlanta, Georgia: Southern Regional Council, 1977. (2D 131124)
- Minskoff, E.H., "Research on Psycholinguistic Training: Critique and Guidelines." Exceptional Children 42 (1975): 136-144.
- Moon, T.C., "A Study of Verbal Behavior Patterns in Primary Grade Classrooms During Science Activities." Journal of Research in Science Teaching 8 (1971): 171-177.
- Morine-Darshimer, G., "How Teachers See Their Pupils." Educational Research Quarterly 3 (Winter 1978-79): 43-53.
- Naron, N.K., "Early Assessment and Remediation of School Learning Problems in a Large Urban Setting: A Realistic Approach." Paper presented at annual meeting of the American Educational Research Association, New York, April 1977.
- National Education Association. Guidelines and Cautions for Considering Criterion-Referenced Tests. Washington, D.C.: National Education Association, 1975.
- National Education Association. Resolution 74-36 (1974).

- National Education Association. Critical Moments in Teaching. Washington, D.C.: National Institute of Education, 1976. (ED 171686)
- National Education Association. The Effective Use of Observation as a Diagnostic Technique: Diagnostic Learning Package 8. Washington, D.C.: National Institute of Education, 1977. (ED 164529)
- National Education Association. Designing Effective Instruction -- A New Adventure in Teaching. Washington, D.C.: National Institute of Education, 1977a. (ED 168999)
- National Education Association. Evaluation of Team Teaching and Children's Continuous Progress. Washington, D.C.: National Institute of Education, 1977b. (ED 166158)
- National Education Association. The Effective Use of Observation as a Diagnostic Technique. Washington, D.C.: National Institute of Education, 1977c. (ED 164529)
- National Education Association. Administration of a Standardized Test. Washington, D.C.: National Institute of Education, 1977d. (ED 164530)
- National Education Association. Parents and Testing. Washington, D.C.: National Education Association, 1979.
- National Education Association. NEA Focus: Standardized Testing. Washington, D.C.: National Education Association, 1979.
- National Education Association. Teachers and Testing. Washington, D.C.: National Education Association, 1979.
- Nearine, R.J., "The Test, the Time, and the Teacher." Measurement and Evaluation in Guidance 2 (Winter 1970): 214-216.
- Newcomer, P., Larsen, S.C., and Hammill, D.D., "A Response to Research on Psycholinguistic Training." Exceptional Children 42 (1975): 144-148.
- Noll, V.H., "Requirements in Educational Measurement for Prospective Teachers." School and Society 82 (September 17, 1955): 88-90.
- Nolte, M.C., "Use and Misuse of Tests in Education: Legal Implications." Educational Horizons (Fall 1975): 10-16.
- NSBA Research Report 1977-1. Standardized Achievement Testing. Washington, D.C.: National School Board Association, 1977, p. 53.
- Nunnally, J.C., Psychometric Theory. New York: McGraw-Hill, 1967.
- Nurss, J. and McGauvran, M.E., Metropolitan Readiness Tests (K to 1.3). "Teacher's Manual Part II: Interpretation and Use of Tests Results." New York: Harcourt, Brace, Jovanovich, 1976.

- Olejnik, S.F., "Standardized Achievement Programs Viewed from the Perspective of Non-Measurement Specialist." Paper presented at annual meeting of the National Council on Measurement in Education, San Francisco, April 1979.
- Openshaw, K., "A Failure of the Minnesota Teacher Attitude Inventory to Relate to Teacher Behavior." Journal of Teacher Education 18 (1967): 233-239.
- O'Regan, M., Airasian, P., and Madaus, G., "The Use of Standardized Test Information by Irish Teachers." Paper presented at annual meeting of the National Council on Measurement and Education, San Francisco, April 1979.
- Page, W. and Carlson, K., "The Process of Observing Oral Reading Scores." Reading Horizons 15 (1975): 147-150.
- Pedulla, J.J., Airasian, P., Madaus, G., and Kellaghan, T., "Proportion and Direction of Teacher Rating Changes of Pupils' Progress Attributable to Standardized Test Information." Journal of Educational Psychology 69 (1977): 702-709.
- Perrone, V., The Abuses of Standardized Testing. Bloomington, Indiana: Phi Delta Kappa Educational Foundation, 1977.
- Popham, W.J., "Objectives and Instruction." In W.J. Popham, et al., Institutional Objectives. Skokie, Illinois: Rand McNally, 1972, 32-52.
- Popham, W.J. and Baker, E.L., Systematic Instruction. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.
- Popham, W.J. and Husek, T.R., "Implications of Criterion-Referenced Measurement." Journal of Educational Measurement 6 (Spring 1964): 1-9.
- Prescott, G.A., Balow, I.H., Hogan, T.P., and Farr, R.C., Metropolitan Achievement Tests (K to 12.9). "Teacher's Manual for Administering and Interpreting," New York: Harcourt, Brace, Jovanovich, 1979.
- Proger, B.B., Cross, L.H., and Burger, R.M., "Construct Validation of Standardized Tests in Special Education: A Framework of Reference and Application to ITPA Research." In The First Review of Special Education, Volume 1. L. Mann and D.A. Sabatino (eds.). Philadelphia: JSE Press, 1973.
- Quinto, F. and McKenna, B., Alternatives to Standardized Testing. Washington, D.C.: National Education Association, 1977.
- Resnick, D.P. and Schumacker, C., "The Pittsburgh Research and Measurement Bureau, 1918-1939." Paper presented at annual meeting of American Educational Research Association, Boston, April 1980.
- Roeder, H.H., "Are Today's Teachers Prepared to Use Tests?" Peabody Journal of Education 49 (April 1972): 239-240.
- Rosenshine, B., "The Stability of Teacher Effects Upon Student Achievement." Review of Educational Research 40 (1970): 647-662.

- Rosenshine, B. and Furst, N., "The Use of Direct Observation to Study Teaching." In Second Handbook of Research on Teaching, R. Traverse (ed.). Chicago: Rand McNally, 1973.
- Rosenthal, R. and Jacobson, L., Pygmalion in the Classroom: Teacher Expectations and Pupils' Intellectual Development. New York: Holt, Rinehart and Winston, 1968.
- Rosner, B., "Substance of Training in Measurement and Evaluation in Pre-service Preparation of Teachers." National Council of Measurement in Education, 18th Yearbook, 1961: 51-54.
- Rudman, H.C., "How Good Are Standardized Achievement Tests." National Elementary Principal 44 (November 1964): 32-38.
- Rudman, H.C., "Between Us - An Analysis of the Michigan Accountability Activities." Michigan Elementary Principal (November/December 1970): 9-11, 14-15; (January/February 1971): 13-14.
- Rudman, H.C., The Michigan Educational Assessment Program: A Technical Analysis of the Michigan Assessment of Basic Skills. East Lansing: Michigan Association of the Professors of Educational Administration - Task Force on Educational Assessment and Accountability, 1973: 44.
- Rudman, H.C., "Between Us - An Analysis of the Michigan Educational Assessment Program." Michigan Elementary Principal (February 1974): 8, 10; (April 1974): 11-12; (June 1974): 10-11, 14; (December 1974): 6, 11; (February 1975): 8, 13.
- Rudman, H.C., "How to Use Standardized Tests for Decision Making." National Elementary Principal 55 (March/April 1976): 59-64.
- Rudman, H.C., "The Standardized Test Flap." Phi Delta Kappan 59 (November 1977): 179-185.
- Rudman, H.C., et al., "Integrating Assessment With Instruction: A Review (1922-1980)." East Lansing, Michigan: Institute for Research on Teaching, Research Series No. 75, 1980.
- Russo, N.A., "The Effects of Student Characteristics, Educational Beliefs and Instructional Task on Teachers' Preinstructional Decisions in Reading and Math." Paper presented at annual meeting of the American Educational Research Association, Toronto, March 1978.
- Salvia, J. and Ysseldyke, J.E., Assessment in Special and Remedial Education. Boston: Houghton-Mifflin, 1978.
- Scannell, D.P., Haugh, O.M., Schild, A.H., and Ulmer, G., Tests of Achievement and Proficiency. "Teacher's Guide, Tests of Achievement and Proficiency, Form T" (Examination Kit Edition) Boston: Houghton-Mifflin, 1978.
- Schultz, J. and Florio, S., "Stop and Freeze: The Physical Space in a Kindergarten/First Grade Classroom." Anthropology and Educational Quarterly, 1979.

- Shanker, A. as cited in Achievement Testing and Basic Skills: Conference Proceedings, edited by T.M. Tomlinson and M.E. Treacy. Washington, D.C.: U.S. Department of Health, Education, and Welfare and National Institute on Education, 1979: 11.
- Shavelson, R.J., "A Model of Teacher Decision Making." Paper presented at annual meeting of the American Educational Research Association, Toronto, March 1978.
- Shavelson, R.J. and Borko, H., "Research in Teachers' Decisions in Planning Instruction." Educational Horizons (Summer 1979): 183-189.
- Shavelson, R.J., Cadwell, J., and Izu, T., "Teachers' Sensitivity to the Reliability of Information in Making Pedagogical Decisions." American Educational Research Journal 14 (Spring 1977): 83-97.
- Shavelson, R.J. and Dempsey-Atwood, N., "Generalizability of Measures of Teaching Behavior." Review of Educational Research 46 (1976): 553-611.
- Shields, P.H. and Neal, A.W., "The 'Action Model' for Staff Development in a Secondary School." Paper presented at annual meeting of the International Reading Association, Houston, May 1978. (ED 159623)
- Shulman, L., Test Design: A View from Practice. Occasional Paper No. 8. East Lansing, Michigan: Institute for Research on Teaching, 1978.
- Skager, R.W., "Student Entry Skills and Evaluation of Instruction Programs: A Case Study." A paper presented at annual meeting of the American Educational Research Association, 1969.
- Smith, F.J. and Luginbuhl, J.E.R., "Inspecting Expectancy: Some Laboratory Results of Relevance for Teacher Training." Journal of Educational Psychology 68 (1976): 265-272.
- Snell, M.E., Taylor, K.G., and Thompson, M.S., "The Active Response Inservice Models: A Method of Providing Inservice to Educators of the Severely Handicapped." In C.M. Nelson, editor, Field Based Teacher Training: Applications in Special Education. Minneapolis: Department of Psychoeducational Studies, University of Minnesota, 1978.
- Snell, M., Thompson, M.S., and Taylor, K., "Providing Inservice to Educators of the Severely Handicapped: The Active Response Inservice Training Model." Education and Training of the Mentally Retarded (February 1979): 25-33.
- Sorotzkin, F., Fleming, E., and Anttonen, R., "Teacher Knowledge of Standardized Test Information and Its Effects on Pupil I.Q. and Achievement." Journal of Experimental Education 43 (Fall 1974): 79-85.
- Spache, G.D., Diagnostic Reading Scales. Monterey: CTB/McGraw-Hill, 1963.
- Sproull, L. and Zubrow, D., "Standardized Testing in the Educational Organization Administrative Performance Information System." Paper presented at annual meeting of American Educational Research Association, Boston, April 1980.

Stanley, J.C., "The Predictive Value of the SAT for Brilliant Seventh and Eighth Grades." College Board Review 107 (Winter 1977 - 78): 31-37.

Stetz, F., Report of a Market Survey to Sample Opinions of Sales Representatives and Users Concerning the 1973 Stanford Achievement Test. Market Research Report No. 10. New York: The Psychological Corporation, 1977: 25.

Stetz, F.P. and Beck, M.D., "A Survey of Opinions Concerning Users of Educational Tests." Paper presented to the National Council on Measurement in Education, Toronto, 1978.

Stetz, F. and Beck, M.D., "Comments from the Classroom: Teachers' and Students' Opinions of Achievement Tests." Paper presented at annual meeting of National Council on Measurement in Education, San Francisco, April 1979.

Swanson, D.H. and Denton, J.J., "A Comparison of Remediation Systems Affecting Achievement and Retention in Mastery Learning," 1976. (ED 131037)

Taba, H., Curriculum Development, Theory and Practice. New York: Harcourt, Brace and World, 1962.

Test Service Bulletin. New York: The Psychological Corporation. Helpful information for users includes: "What Is an Aptitude" (No. 36); "How Effective Are Your Tests" (No. 37); "Expectancy Tables -- A Way of Interpreting Test Validity" (No. 38); "Norms Must Be Relevant" (No. 39); "The Three-Legged Coefficient" (No. 40); "Human Resources and the Aptitude Inventory" (No. 41); "Does Testing Cost Too Much?" (No. 42); "The Search for Talent" (No. 43); "Reliability and Confidence" (No. 44); "Better Than Chance" (No. 45); "The Correction for Guessing" (No. 46); "Cross-Validation" (No. 47); "Methods of Expressing Test Scores" (No. 48); "The DAT - A Seven-Year Follow-Up" (No. 49); "Aptitude, Intelligence, and Achievement" (No. 51); "Watch Your Weights" (No. 52); "Comparability vs. Equivalence of Test Scores" (No. 53); "On Telling Parents About Test Results" (No. 54); "The Identification of the Gifted" (No. 55); "Double-Entry Expectancy Tables" (No. 56); "Testing Job Applicants from Disadvantaged Groups" (No. 57); "Local Norms-When and Why" (No. 58); "Restriction of Range: Questions and Answers" (No. 59).

Test Service Notebook. New York: The Psychological Corporation. Helpful materials for users include: "A Glossary of Measurement Terms" (No. 13); "Testing in the Secondary School" (No. 20); "Accountability in Education and Associated Measurement Problems" (No. 33); "Some Things Parents Should Know About Testing" (No. 34); "Testing: Bond or Barrier Between Pupil and Teacher?" (No. 82); "Selection and Provision of Testing Materials" (No. 99); "Stanines and Their Computation for Local Use" (No. 123); "How a Standardized Achievement Test Is Built" (No. 125); "Innovation in the Assessment of Individual Differences" (No. 130); "On Telling Parents About Test Results" (No. 154).

Thorndike, R.L. and Hagen, E., Cognitive Abilities Test. "Examiner's Manual, Form 3" (Examination Kit Edition). Boston: Houghton Mifflin, 1978.

Tikunoff, W.J. and Ward, B.A., "Problems in Conducting Naturalistic Research: An Analysis of a Case Study of Classroom Socialization." Paper presented at annual meeting of American Educational Research Association, Toronto, March 1978.

- Tolor, A., Scarpetti, W.L., and Lane, P.A., "Teachers' Attitudes Toward Children's Behavior Revisited." Journal of Educational Psychology 58 (1967): 175-180.
- Tomlinson, T.M. and Treacy, M.E., editors. Achievement Testing and Basic Skills: Conference Proceedings. Washington, D.C.: U.S. Department of Health, Education and Welfare and National Institute of Education, 1979: 25-26, 28-29.
- Tractenberg, P.L. and Jacoby, E., Pupil Testing: A Legal View. Washington, D.C.: U.S. Department of Health, Education and Welfare and National Institute of Education, 1974.
- Tversky, A. and Kahneman, D., "Judgment Under Uncertainty: Heuristics and Biases." Science 185 (1974): 1124-1131.
- Tyler, R.W., Basic Principles in Curriculum and Instruction. Chicago, Illinois: University of Chicago Press, 1950.
- Tyler, R.W. and White, S.H., Testing, Teaching and Learning: Report of a Conference of Research on Testing. Washington, D.C.: National Institute of Education, 1979.
- United States Department of Defense. Basic Skills Assessment Program (filmstrip and cassette), 1979.
- Walker, C.B., Criterion-Referenced Test Evaluation Handbook, Los Angeles, CA: UCLA Center for the Study of Evaluation, 1977.
- Washburne, C.W., "Educational Measurements as a Key to Individualizing Instruction and Promotions." Journal of Educational Research 5 (1922): 195-206.
- Waugh, R.P., "ITPA: Ballast or Bonanza for the School Psychologists." Journal of School Psychology 13 (1975): 201-208.
- Weckstein, P., "Legal Challenges to Educational Testing Practices." Inequality in Education 15 (November 1973): 92-101.
- Weiner, C., Basic Skills Assessment (7-12) "Basic Skills Assessment User Guide". Menlo Park, California: Addison-Wesley, 1978.
- Wexley, K.N. and Thornton, C.L., "Effect of Verbal Feedback of Test Results Upon Learning." Journal of Educational Research 66 (November 1972): 119-121.
- Wise, A.E., "The Hyper-Rationalization of American Education." Educational Leadership (February 1978): 354-361.
- Yeany, R.H., et al., "The Effects of Diagnostic Prescriptive Instruction and Locus of Control on the Achievement and Attitudes of University Students." Paper presented at annual meeting of The National Association for Research in Science Teaching, Atlanta, March 1979. (ED 171579)
- Yeh, J.P., "Test Use in Schools." Washington, D.C.: U.S. Department of Health, Education and Welfare and National Institute of Education, 1978: 9-42.

Ysseldyke, J.E., "Diagnostic-Prescriptive Teaching: The Search for Aptitude-Treatment Interactions." In The First Review of Special Education, Vol. I, L. Mann and D.A. Sabatino (eds.). Philadelphia: JSE Press, 1973.

Zahorik, J.A., "Classroom Feedback Behavior of Teachers." Journal of Educational Research 62 (1968): 147-150.

LIST OF PARTICIPANTS

NATIONAL INSTITUTE OF EDUCATION
INVITATIONAL CONFERENCE ON TEST USE

June 2, 1980
Washington, D.C.

Lawrence M. Rudner, Chair
National Institute of Education
1200 19th Street, N.W.
Washington, D.C. 20208

Jeff Schiller
Associate Director
Testing, Assessment and Evaluation
National Institute of Education
1200 19th Street, N.W.
Washington, D.C. 20208

PRESENTORS

Michael Beck
The Psychological Corporation
757 Third Avenue
New York, New York 10017

Walter Haney
Huron Institute
123 Mt. Auburn Street
Cambridge, Massachusetts 02138

Mary Kennedy
Huron Institute
123 Mt. Auburn Street
Cambridge, Massachusetts 02138

Herbert Rudman
Department of Administration and
Higher Education
College of Education
Erickson Hall
Michigan State University
East Lansing, Michigan 48824

Leslie Salmon-Cox
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

James Ward
American Federation of Teachers
11 Dupont Circle, N.W.
Washington, D.C. 20036

Jennie Yeh
Center for the Study of Evaluation
145 Moore Hall
University of California
Los Angeles, California 90024

CONFERENCE PARTICIPANTS

Cynthia Almeida
 Clyde C. Aveilhe
 Robert Bigelow
 Brenda Biles
 Gerry Bogatz
 Mary Lynn Bourgue
 William J. Brown
 Jerome H. Bruns
 M.M. Coleman
 John Crowder
 Lois-ellin Datta
 Charles D. Duffy
 Mary Fruen
 Pamela Getson
 Jewell Gould
 Dennis Gray
 William Griffith
 Carl Guerriero
 Enid Herndon
 Holly Hexter
 Dennis Holmes
 Robert Jones
 Ann Kahn
 A. Brian Klein
 Stephen Koffler
 Edward L. Loh
 Carlyle E. Maw
 Bernard McKenna
 Mary Ann Milsap
 Karen Mitchell
 Lewis Pike
 Joseph Psotka
 Francis Quinto
 Philip R. Rever
 William J. Russell
 Judy Shoemaker
 Pam Splaine
 Richard Stiggins
 Lucinda Tursman
 Janet Wall
 Kathi Yohn

District of Columbia Public Schools
 Educational Testing Service
 Delaware State Department of Public Instruction
 American Federation of Teachers
 Educational Testing Service
 University of Maryland
 North Carolina State Department of Public Instruction
 Falls Church (VA) Public Schools
 Scott, Foresman Publishing Company
 Arlington (VA) Public Schools
 National Institute of Education
 West Virginia State Department of Education
 Department of Health and Human Services
 Gallaudet College
 American Federation of Teachers
 Council for Basic Education
 Fairfax (VA) County Public Schools
 Pennsylvania State Department of Education
 National Institute of Education
 Higher Education Daily
 Department of Defense Dependent Schools
 Association of American Medical Colleges
 National Congress of Parents and Teachers
 Charles County (MD) Public Schools
 New Jersey State Department of Education
 Prince George's County (MD) Public Schools
 Office for Civil Rights
 National Education Association
 National Institute of Education
 National Institute of Education
 National Institute of Education
 National Institute of Education
 National Education Association
 American College Testing Program
 National Council on Measurement in Education
 National Institute of Education
 Montgomery County (MD) Public Schools
 Northwest Regional Educational Laboratory
 American Association of School Administrators
 Department of Defense Dependent Schools
 National Institute of Education